

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 1, 2002, 20:06:13 ; Search time 1485.96 Seconds
(without alignments)
8404.240 Million cell updates/sec

Title: US-09-586-106-62

Perfect score: 757

Sequence: 1 gaaagaggtgtgaagctc.....atcaagattcacaaggt 757

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: 1: gb_ba: 2: gb_htg: 3: gb_in: 4: gb_om: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_sts: 12: gb_sy: 13: gb_un: 14: gb_vl: 15: em_ba: 16: em_fun: 17: em_hum: 18: em_in: 19: em_om: 20: em_or: 21: em_ov: 22: em_pat: 23: em_ph: 24: em_pl: 25: em_ro: 26: em_sts: 27: em_sy: 28: em_un: 29: em_vl: 30: em_htgo_hum: 31: em_htgo_inv: 32: em_htgo_rod: 33: em_htg_hum: 34: em_htg_inv: 35: em_htg_rod: 36: em_htg_other:

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	571.4	75.5	65979	8	AF254799	AF254799 Hordeum v
2	389.4	51.4	83646	8	AB005248	AB005248 Arabidops
3	381.6	50.4	49981	6	AX059550	AX059550 Sequence
4	381.6	50.4	54708	8	AC007209	AC007209 Arabidops
5	381.6	50.4	63292	8	F14616	AF147260 Arabidops
6	381.6	50.4	198176	8	ATCHRIV17	AL161505 Arabidops
7	380.8	50.3	95685	8	AC007187	AC007187 Arabidops
8	380.8	50.2	40480	6	AX059456	AX059456 Sequence
9	380	50.2	76911	2	AB016828	AC016828 Arabidops
10	380	50.2	85962	8	AB046431	AB046431 Arabidops
11	380	50.2	87814	8	AB046426	AB046426 Arabidops
12	380	50.2	95463	8	F1809	AF1809 Arabidops
13	380	50.2	118986	8	AC007534	AC007534 Arabidops
14	378.4	50.0	54573	8	T619	AB026642 Arabidops
15	378.4	50.0	77287	8	AB026642	AB026642 Arabidops
16	378.4	50.0	183181	8	ATCHRIV19	AF147261 Arabidops
17	376.8	49.8	18813	8	T9E19	AB028613 Arabidops
18	376.8	49.8	35313	8	AB028613	AB028613 Arabidops
19	376.8	49.8	51248	8	AB046437	AB046437 Arabidops
20	376.8	49.8	67701	8	AC007155	AC007155 Arabidops
21	376.8	49.8	77204	8	AB046438	AB046438 Arabidops
22	376.8	49.8	108227	8	ATF/P3	AL138653 Arabidops
23	376.8	49.8	121501	8	AC069557	AC069557 Genomic S
24	375.2	49.6	37010	6	AX059491	AX059491 Sequence
25	375.2	49.6	38519	6	AX059548	AX059548 Sequence
26	375.2	49.6	42112	6	AX059490	AX059490 Sequence
27	375.2	49.6	47347	6	AX059547	AX059547 Sequence
28	375.2	49.6	48422	6	AX059509	AX059509 Sequence
29	375.2	49.6	48422	6	AX059509	AX059509 Sequence
30	375.2	49.6	79122	8	ATAC009261	AF296829 Arabidops
31	375.2	49.6	88126	8	F3D18	AF147261 Arabidops
32	375.2	49.6	94905	8	F2112	AF147261 Arabidops
33	375.2	49.6	94905	8	AC019013	AL161504 Arabidops
34	375.2	49.6	114056	8	ATCHRIV16	AL161504 Arabidops
35	375.2	49.6	196766	8	ATCHRIV16	AL161504 Arabidops
36	375.2	49.6	196766	8	ATCHRIV16	AL161504 Arabidops
37	372	49.1	107487	2	AC087569	AC087569 Arabidops
38	371.2	49.0	39104	6	AX059494	AX059494 Sequence
39	371.2	49.0	87219	8	AP002054	AP002054 Arabidops
40	371.2	49.0	129021	8	ATT32A11	AL138653 Arabidops
41	370.4	48.9	36096	6	AX059471	AX059471 Sequence
42	370.4	48.9	38243	6	AX059470	AX059470 Sequence
43	370.4	48.9	103931	8	AC020646	AC020646 Genomic S
44	370.4	48.9	108720	8	AB046436	AB046436 Arabidops
45	370.4	48.9	134499	8	AC007505	AC007505 Arabidops

ALIGNMENTS

RESULT	1	PLN	01-AUG-2000
AF254799/c	AF254799	65979 bp	DNA
LOCUS	AF254799		
DEFINITION	Hordeum vulgare (onoplact intrinsic protein 1 (TIP1), tonoplast intrinsic protein 2 (TIP2), and Kar1 (Kar1) genes, complete cds.		
ACCESSION	AF254799		
VERSION	AF254799.1	GI:9623334	
KEYWORDS			
SOURCE	Barley.		
ORGANISM	Hordeum vulgare		
REFERENCE	Shirasu K., Schulman A.H., Lahaye T. and Schulze-Lefert P.		
AUTHORS	A contigous 66-kb barley DNA sequence provides evidence for reversible genome expansion		
TITLE	Genome Res. 10 (7), 908-915 (2000)		
JOURNAL	20359834		
MEDLINE			
REFERENCE	2 (bases 1 to 65979)		

AUTHORS Shirasu, K., Schulman, A.H., Lahaye, T. and Schulze-Lefert, P.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2000) The Sainsbury Laboratory, John Innes
Centre, Colney Lane, Norwich NR4 7UH, UK

FEATURES
Source Location/Qualifiers
1. 65979

/organism="Hordeum vulgare"
/db_xref="taxon:4513"
/chromosome="2"
/map="2HL: MNG87"
402..5501

misc_feature
/note="similar to BAGY-2 polyprotein"
5608..6244

LTR
/note="BAGY-2 LTR-1"
6245..11204

LTR
/note="Sukula LTR-1"
11205..11757

LTR
/note="BAGY-2 LTR-1"
11758..13280

LTR
/note="BAGY-2 LTR-2"
13281..13627

LTR
/note="BAGY-2 LTR-1"
14378..14966

LTR
/note="Sabra LTR-1"
14967..16796

LTR
/note="BARE-1 LTR-1"
16797..17590

LTR
/note="Sabra LTR-1"
19160..20144

LTR
/note="Sabra LTR-2"
20145..24048

LTR
/note="Sukula LTR-2"
24049..23865

LTR
/note="BARE-1 LTR-2"
23866..27570

LTR
/note="Sukula LTR-2"
27571..28211

LTR
/note="Sabra LTR-2"
31781..33400

LTR
/note="Sabra LTR-3"
33897..34585

LTR
/note="Nikita LTR-1"
34586..36407

LTR
/note="BARE-1 LTR-3"
36418..38442

LTR
/note="BARE-1 LTR-4"
38443..40171

LTR
/note="Nikita LTR-1"
40261..41734

LTR
/note="similar to lnp2"
40261..41734

misc_feature
/note="similar to lnp2"
40261..41734

gene
/product="lonoplast intrinsic protein 1"
444020..>44979

gene
/gene="TIP1"
444020..44146,44233..44483,44608..44979

gene
/product="lonoplast intrinsic protein 1"
444020..44146,44233..44483,44608..44979

gene
/product="lonoplast intrinsic protein 1"
444020..44146,44233..44483,44608..44979

gene
/product="lonoplast intrinsic protein 1"
444020..44146,44233..44483,44608..44979

gene
/product="lonoplast intrinsic protein 1"
444020..44146,44233..44483,44608..44979

gene
/product="lonoplast intrinsic protein 1"
444020..44146,44233..44483,44608..44979

gene
/product="lonoplast intrinsic protein 1"
444020..44146,44233..44483,44608..44979

gene
/product="lonoplast intrinsic protein 1"
444020..44146,44233..44483,44608..44979

/codon_start=1
/product="lonoplast intrinsic protein 2"
/protein_id="AAP90122.1"
/db_xref="GI:9623336"
/translation="MASNLRYHLKHCSPSPSLRSYFAEFTSTLEFVAVGSAISARM
LTPDVTSNASSSLAVATAVAOSFGLEAAVFAADVSGGVNPAVFAFAGIGVPTAI
FYWTQILGSLTACLVHFLSAGQAVPTTRIAVMTGEGASIVGVTFWVTVYVHA
GDPRGGRKGLATSLGALVGLVGTACVLAAGSLTGSNMPARSFPAVSGDFKNO
AVYAGPMGAVALVHQLNLPSPAPRPLPHEBSHSGVETVV"
47279..47443

repeat_region
/rpt_family="Stowaway"
/rpt_type="Inverted"
50003..53152

misc_feature
/note="similar to Tail"
join(<56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/product="Rar1"
56694..>60254

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

CDS
/gene="Rar1"
/product="Rar1"
56694..>60254

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

gene
/gene="Rar1"
join(56694..56822,58480..58482,58563..58645,59572..59738,
59853..60076,60162..>60254)

Oy 421 tggcatatctctactcttgatgaaatattcgcagagtttcatgagatcttccg 480
 ||||||| ||| ||||||||| ||| ||||||||| ||| |||
 Db 3028 TGGCTATATCTCTGACTTTTGAAAGATTTATGAGTTTCATGATGATTTCCG 2969
 Oy 481 ttacaggctctctcttgatgagatgctcgaaccttgatgagatcttcaagatgta 540
 ||||||| ||| ||||||||| ||| ||||||||| ||| |||
 Db 2968 TTTACGGTTCTTCTTGGAGACTACTCCTGACCACTTGATCGATTTTACAAAGATGCA 2909
 Oy 541 aagaccacatctcttgatgagatgagagagatgagacatcttgatgagatgagacacg 600
 ||||||| ||| ||||||||| ||| ||||||||| ||| |||
 Db 2908 AAGATACCACTCTGCTGCTGACTGAGAGAGATGCTCCTTCATGATGATGATGATGATG 2849
 Oy 601 tcttagagacataaattcttgaagagatgagatgagatgagatgagatgagatgagac 660
 ||||||| ||| ||||||||| ||| ||||||||| ||| |||
 Db 2848 TCTTAGAGACCAAAATCTCTGCTGAGAGAGATGATGATGATGATGATGATGATGATG 2789
 Oy 661 tgcagaaatgcatgacccacagatgagatgagatgagatgagatgagatgagatgagatg 720
 ||| ||||||| ||| ||||||||| ||| ||||||||| ||| |||
 Db 2788 TTGAGAAATGCTTGGCCCTGATGATGATGATGATGATGATGATGATGATGATGATG 2729
 Oy 721 gttctatagaagtttatacaagattcaacaagt 757
 ||||||| ||| ||||||||| ||| ||||||||| ||| |||
 Db 2728 GTTCTATAGAGGTTCTATTAGAGACTTCTTATGAT 2692

RESULT 2
 AB005248 83646 bp DNA PLN 27-DEC-2000
 LOCUS Arabidopsis thaliana genomic DNA, chromosome 5, pl clone:MX110.
 DEFINITION AB005248 BA000015
 ACCESSION AB005248.1 GI:2264320
 VERSION
 KEYWORDS
 SOURCE

ORGANISM Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui pl clone:MX110.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 REFERENCE 1 (sites)
 AUTHORS Sato,S., Kotani,H., Nakamura,Y., Kaneko,T., Asamizu,E., Fukami,M., Miyajima,N. and Tabata,S.
 TITLE Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence features of the 1.6 Mb regions covered by twenty physically assigned pl clones
 JOURNAL DNA Res. 4 (3), 215-230 (1997)
 MEDLINE 97471969
 REFERENCE 2 (bases 1 to 83646)
 AUTHORS Nakamura,Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUL-1997) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)
 COMMENT Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/agd.graph.cgi?c=MX110
 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/grail-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremml.zool.iastate.edu/cgi-bin/sp.cgi).
 Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/).
 This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions.

FEATURES
source

The 5' clone is MS17 and the 3' clone is MB18.

Location/Qualifiers
 1..83646
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="MX110"
 /clone_lib="Mitsui pl"
 join(1791..2912,3016..3318,3584..4660)
 /note="gene_id:MX110.1"
 /codon_start=1
 /evidence=not-experimental
 /product="disease resistance protein-like"
 /protein_id="BAB09346.1"
 /db_xref="GI:9758812"
 /translation="MTERKADVSNMLNNSPRDPTGMDAMKMEISLCLDSD
 /VRMIGWPGSIGRTTARVLYSQSFSENFELSIPMGKIKELMTTRPVCSPEAKIO
 LOKOFLSQIINRDMELHLYGADRLNDKVLIVLSDIDSIDLDIAKETRMRGVT
 SRITITODOKLRKAGINHIYKVEFPAYEAYOMFCMYARGQNPNDGFELEMEVT
 KLIGHPLGLRYVSGHFGMSRHEWVNALPRLKRLDASIOSILKFSYDALCEBDKL
 FLHICLFPNOMVEVEDYALSFIDVGGFHLIAEKSLINLKFSTNCTLEMHNL
 VVDKDIVNRKPGHOSICPRKROFLIDARDICEVLTNTGTRNVYGIFFLEVRNLSC
 LNI SERAFEGMSNLKFLRFHPDYDESDRLYLPGLNLNLPKRLIEMSRPMTCLPS
 NECTKYLVEIRMKNSKLQMLWQNOPLGKLMKMDISEKHLEKLPDLSTATNLELLH
 SCSISLVELPSSIGRLKRLMLSLRGSRLKLEALPINLESLDYLDLDCILKFPK
 ISGNIKDLTKRTAIKEVPSITKSMHLEKSENLKELPHALDIITLYINDTE
 MOEIPOMWKKISHILOTILGEGKRLVITIPOLSDISQIVTNCESLERLNSFQWPE
 RFLMFLNCRKLNNAREPQTQSTTHAILLPREPANPTYRANGSSIMVNLNRPSTT
 LRFRKACVLLVKRLINDKEADRRRTTVIRIENKIGCVWRYRFRVPTLDEHLL
 TQIIEADVTSNKLSEFEDIKDNNQAVKRCGLQI"
 complement(join(5504..5620,5688..5736,5836..5905,
 5990..6122,6206..6250,6374..6688))
 /note="gene_id:MX110.2"
 pit||T03594
 similar to unknown protein"
 /codon_start=1
 /evidence=not-experimental
 /protein_id="BAB09347.1"
 /db_xref="GI:9758813"
 /translation="MMASSSLSPFCLSEFKMPIALITNSPTTIPYLNKKGOTRLH
 KLPGLKMEYIERKSKREPNPLPVFVGSYHAACWAMENMLPFSSGSPSYAVSL
 LGGSEDEPLGYAGTLQTHASDIAPFISNCGSPPVVGHSGGLIYOYLVANV
 KRLIGENAPPELISGAVWCSPVSGNSGIVLRLTSKVAAPKYLISLAAGFPQKSI
 PLCRETFEQAMDOLVKRYNLAR"
 join(7788..7835,8043..8777,9032..9157,9205..9901,
 10085..12958)
 /note="gene_id:MX110.3"
 /codon_start=1
 /evidence=not-experimental
 /product="non-LTR retroelement reverse transcriptase-like protein"
 complement(13525..13827)
 /note="gene_id:MX110.4"
 pit||T02618
 similar to unknown protein"
 /codon_start=1
 /evidence=not-experimental
 /protein_id="BAB09348.1"
 /db_xref="GI:9758814"
 /translation="MKDYILIVLYSLIKMTKRSQPLDNCSSSETYPOLDKLNKHNDG
 LVLLMVQSQWRCQMKRHEIYVDLSATIEPIHIAVYVHOLPYDITLIFPCLFS"
 join(18792..18841,19220..19412,19494..19579,19875..19959,
 20464..20607,20697..20783,20894..21006,21097..21162,
 21216..21308,21350..21509)
 /note="gene_id:MX110.5"
 unknown protein"
 /codon_start=1
 /evidence=not-experimental
 /protein_id="BAB09349.1"
 /db_xref="GI:9758815"
 /translation="MMSPKQISDRGSSHFRRPTFOIHLIGNFRIWYSWRYLNL

FEATURES		Address all correspondence to: atetigr.org.
source	Location/Qualifiers	
misc_feature	1..54708	/organism="Arabidopsis thaliana"
	/db_xref="taxon:3702"	/chromosome="11"
misc_feature	complement(1..434)	/note="Sequence from clone F15011"
	complement(435..54507)	/note="Sequence from clone F1404"
gene	complement(435..1687)	/note="F1404.9"
	/note="F1404.9"	/pseudo
mRNA	complement(<4744..>5949)	/gene="At2g13350"
	complement(<4744..>5949)	/gene="At2g13350"
gene	/note="F1404.8; identical to hypothetical protein GB:AAD20424"	
	complement(4744..5949)	/gene="At2g13350"
CDS	/note="hypothetical protein"	/codon_start=1
	/protein_id="AAD28670.1"	/db_xref="GI:4733989"
mRNA	/translation="MSFLAPCOILELNIISAOELAPARCKTYAIAWIDPERLTLRVDNMGSTPMNCKEYFRIDBEALYDXTSTVYEITVILHFKDTHVGVOLISDLVS	
	PSAMRVTLEVLASGRPHGLNIAVGLDINSQSMPLEEDLHFKNLSKPG	
gene	LRSKSDTSSVDSPPKRVQSVSTNSSEDSOMVVEYEPQRTNPMROT	
	KQDLVGTGPMKPRVTATYTPKRNITIEGTSPMRSPVITTESDLSPASVAAQIAKK	
CDS	ALUGKAESTVTSVGERSEVGLSKTERMOANLEPVLDGVSYPQSDYPTNSNENK	
	SSKPNIEIVRNPMIGAPIQKPSGRNKGSGDGLFCFNCIGIECSYVCGSSGQK	
mRNA	AAKKKKKK	
	complement(join(7452..7773,7852..8062,8182..8548,	
gene	8632..9180,9454..9489))	
	/gene="At2g13360"	
CDS	complement(7452..9489)	
	/note="F1404.7"	
mRNA	complement(join(7636..7773,7852..8062,8182..8548,	
	8632..9121))	
gene	/gene="At2g13360"	
	/codon_start=1	
repeat_region	/product="alanine-glyoxylate aminotransferase"	
	/protein_id="AAD28669.1"	
repeat_region	/db_xref="GI:4733989"	
	/translation="MDYMGGRPHNLFGVGPVNIPEPVIRAMRNEDVRSPIALIT	
misc_feature	KTLLEPVKKIKFTSTSTPFLPTTGANESALNTLSPEDRIVSELIQGSFLMIDQ	
	OKRLNPNVVEEDMGOGANLOVLAKLSQDENHTIKALCIHNENATGTGIVASIKRA	
gene	TLLDHYKHPALLLVGVSSICALDFPMDEGVVALTGSKALSLPTGLGIVASIKRA	
	LEATKTSKSLKVEFDNDYLFKEKLTSTWYPSIOLGLRALSLDLPFEGEENISIA	
repeat_region	RARLGKATRLAVFAMGLKNCOKKEIMSNIVYAVWPHIDSESLVRAMQVYENISA	
	GGLNKVAVKVFRIHGLGNVNEQLLGLCLAGVEMILKDVGYPVVMSGVAAASTYIQH	
repeat_region	HTLIPSRF"	
	complement(9514..9610)	
repeat_region	/rpt_family="(TAAA)n"	
	complement(join(12400..13193,13358..13451,13543..13896,	
mRNA	13995..14116,14264..14441,14540..14695,14939..15184,	
	15294..15374,15643..15822,15902..15988,16358..16468,	
gene	16558..16800,17381..17587,17747..17810,17888..18012,	
	18110..18203,18561..18628,18716..18874,19158..19319,	
gene	19451..19538,19652..19793,20166..20229,20559..20610,	
	20779..20882,21011..21070,21364..21447,21558..21782,	
gene	22156..22575,22795..22862,23018..>23762))	
	/gene="At2g13370"	
repeat_region	complement(12400..>23762)	
	/gene="At2g13370"	
repeat_region	/note="F1404.6"	
	12464..12500	
repeat_region	/rpt_family="(TA)n"	

CDS

complement(join(12820..13193,13358..13451,13543..13896,
13995..14116,14264..14441,14540..14695,14939..15184,
15294..15374,15643..15822,15902..15988,16358..16468,
16558..16800,17381..17587,17747..17810,17888..18012,
18110..18203,18561..18628,18716..18874,19158..19319,
19451..19538,19652..19793,20166..20229,20559..20610,
20779..20882,21011..21070,21364..21447,21558..21782,
22156..22575,22795..22862,23018..>23762))

/gene="At2g13370"

/codon_start=1

/product="SNF2 family (CHD1 subfamily) chromodomain protein"

/protein_id="AAD28668.1"

/db_xref="GI:4733988"

Query Match	50.4%; Score 381.6; DB 8; Length 54708;	
	Best Local Similarity 69.0%; Pred. No. 9,4e-89;	
Matches 522; Conservativity	0; Mismatches 234; Indels 0; Gaps 0;	
Db	38251 AAAAGAAATTTTGAAGATGCTTGAATGCTGATCATCTACCCATATCTGATGACTTG 38310	
Qy	2 aagaagagtgtaagctcgtatgaagtatatcatcatcagtgctcgaagtgatg 61	
Qy	62 ggtgagtcgagtcagtcctcgaagagagagagagagagagagagagagagag 121	
Db	38311 GGTTCCTCCAGTGCATTCGCTCCCTAAAGAGGATGATACCTGTTCTCAAAATGTA 38370	
Qy	122 ggaatgaatgatcccgagagagattatcactgagctatgaagtgatcagattccgaa 181	


```

repeat_region /note="F14G16"
1528..1684
repeat_region /note="F14G16"
1561..1647
repeat_region /note="F14G16"
1646..1907
repeat_region /note="F14G16"
1672..1814
repeat_region /note="F14G16"
1824..2012
repeat_region /note="F14G16"
1901..1977
repeat_region /note="F14G16"
1975..2479
repeat_region /note="F14G16"
1996..2315
repeat_region /note="F14G16"
2213..2530
repeat_region /note="F14G16"
2465..2565
repeat_region /note="F14G16"
2474..2791
repeat_region /note="F14G16"
2793..2826
repeat_region /note="F14G16"
2826..3601
repeat_region /note="F14G16"
3602..4665
repeat_region /note="F14G16"
5136..5243
repeat_region /note="F14G16"
5241..5443
repeat_region /note="F14G16"
5887..6013
repeat_region /note="F14G16"
6002..6401
repeat_region /note="F14G16"
7139..7282
repeat_region /note="F14G16"
7300..7710
repeat_region /note="F14G16"
7701..7747
repeat_region /note="F14G16"
7738..7837
repeat_region /note="F14G16"
7835..8136
repeat_region /note="F14G16"
8114..8270
repeat_region /note="F14G16"
10255..10398
repeat_region /note="T32N15_del_retroTn"
10261..10435
repeat_region /note="T32N15_del_retroTn"
10424..10834
repeat_region /note="T32N15_del_retroTn"
10824..10865
repeat_region /note="T32N15_del_retroTn"
10824..10870
repeat_region /note="T32N15_del_retroTn"
10860..11259
repeat_region /note="T32N15_del_retroTn"
10860..10959
repeat_region /note="T32N15_del_retroTn"
10958..11259
repeat_region /note="T32N15_del_retroTn"
11231..11812
repeat_region /note="T32N15_del_retroTn"
11238..11394
repeat_region /note="T32N15_del_retroTn"
11820..11906
repeat_region /note="T32N15_del_retroTn"
11906..12167
repeat_region /note="T32N15_del_retroTn"

```

```

repeat_region 12162..12238
repeat_region /note="T32N15_del_retroTn"
12237..12741
repeat_region /note="T32N15_del_retroTn"
12735..13052
repeat_region /note="T32N15_del_retroTn"
13053..13086
repeat_region /note="T32N15_del_retroTn"
13087..13862
repeat_region /note="T32N15_del_retroTn"
13864..14927
repeat_region /note="T32N15_del_retroTn"
14920..15027
repeat_region /note="T32N15_del_retroTn"
15024..15226
repeat_region /note="T32N15_del_retroTn"
15224..15350
repeat_region /note="T32N15_del_retroTn"
15340..15739
repeat_region /note="T32N15_del_retroTn"
21183..22195
gene /gene="F14G16.2"
join(21183..21320,21398..21662,21795..21897,21973..22195)
CDS /gene="F14G16.2"

```

Query Match 50.4%; Score 381.6; DB 8; Length 63292;
 Best Local Similarity 69.4%; Pred. No 9.6e-89;
 Matches 519; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

```

QY 3 aaagaagttggaagctccctgatagaaggtatatactatcatgtgtgcatagcattg 62
Db 49198 AAGGAGATTGTGAACCTGTGGAGCAGGATATATCTTCCTGATTTCAGATGAGTTAG 49257
QY 63 gtgaatccgtgataagcttccctgaagaagagagagagagagagagagagagag 122
Db 49258 GTTAGCTAGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 49317
QY 123 gatgaattgataccgcagagagagagagagagagagagagagagagagagagag 182
Db 49318 GACGAGCTGATTCCTACTCGACATCAGAGCCCGGATGATGATGATGATGATGATG 49377
QY 183 ctgaataaagcccaactgaagaagagagagagagagagagagagagagagagagag 242
Db 49378 ATGAAATGCTGCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 49437
QY 243 aggttgcacaaactcaacactctgcttctagaagcgtatctatgattccccaata 302
Db 49438 ATGTTAGCAATATCAATATCAATATCAATATCAATATCAATATCAATATCAATAT 49497
QY 303 ccagttgcacaaactgaagaagagagagagagagagagagagagagagagagagag 362
Db 49498 CCGATTTCATCGAGATGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 49557
QY 363 taagaagcgtatgcttctgcttctatgtaagcagcagcagcagcagcagcagcag 422
Db 49558 TATCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 49617
QY 423 gcataatctcaacttctgtaaaaataatgtagaggttctcaatgagacattccgct 482
Db 49618 TCTATCTTACAGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 49677
QY 483 tacggctctcttctgtagatgctcagcaactgctgtagcttcaagaagtgttaa 542
Db 49678 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 49737
QY 543 gacaccaactctgtcttgaatgagagagagagagagagagagagagagagagagag 602
Db 49738 GAGAAACATCTAGTTTAAATTTGGAGAAATGATGATGATGATGATGATGATGATG 49797
QY 603 ttagacataaattctgaaagagagagagagagagagagagagagagagagagag 662

```



```

repeat_unit      12162..12238
                  /note="T32N15_del_retroTn"
repeat_unit      12237..12741
                  /note="T32N15_del_retroTn"
misc_feature      12310..15735
                  /note="pseudogene, similar to retrovirus-related
                  polypeptides"
repeat_unit      12735..13052
                  /note="T32N15_del_retroTn"
repeat_unit      13053..13086
                  /note="T32N15_del_retroTn"
repeat_unit      13087..13862
                  /note="T32N15_del_retroTn"
repeat_unit      13864..14927
                  /note="T32N15_del_retroTn"
repeat_unit      14920..15027
                  /note="T32N15_del_retroTn"
repeat_unit      15024..15226
                  /note="T32N15_del_retroTn"
repeat_unit      15224..15350
                  /note="T32N15_del_retroTn"
repeat_unit      15340..15739
                  /note="T32N15_del_retroTn"
gene              21183..22195
                  /gene="AT4g07310"
                  /note="contains similarity to Arabidopsis thaliana
                  hypothetical proteins, see GB:AF058826
                  similarity to"
                  /codon_start=1
                  /protein_id="CAB81101.1"
                  /db_xref="GI:7267327"
                  /translation="MESFHPILSLKPSRGKICGRVRFVLSVSSKVMGLILAE
                  EHGMTIEATVGVKMSDHYKDPINCEWVITTFNCEVNSGSVRAITHSFKIFGVDDIV
                  VRLTSPILPAPHYRLASFSTIIDDEIDKSVLDVAGAIYGVGLINTRPKNIGFVDIV
                  LIRKISDNENKVELCLATKKFALDFDHNRYRGGVIVAVLGMMWKIDRYDGPKNRV
                  CTAGTISTVEPDDIPESNEIHEI"
exon              21183..21320
                  /gene="AT4g07310"
                  /number=1
intron            21321..21397
                  /gene="AT4g07310"
                  /number=1
exon              21398..21662
                  /gene="AT4g07310"
                  /number=2
intron            21663..21794
                  /gene="AT4g07310"
                  /number=3
exon              21795..21897
                  /gene="AT4g07310"
                  /number=4
intron            21898..21972
                  /gene="AT4g07310"
                  /number=3
exon              21973..22195
                  /gene="AT4g07310"
                  /number=4
misc_feature      23319..28183
                  /note="pseudogene, similar to retrovirus-related
                  polypeptides"
exon              complement(29065..29454)
                  /gene="AT4g07320"
                  /number=1
gene              29065..31484
                  /gene="AT4g07320"
                  complement(join(29065..29454,29519..29703,30603..30969,
Query Match      50.4%; Score 381.6; DB 8; Length 198176;
Best Local Similarity 69.4%; Pred. NO. 11e-88;
Matches 519; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

```

```

QY      3 aaagaggttggaagcctccgagatgaagratatcatatcttgcataagcattgg 62
        || || || || || || || || || || || || || || || || || || || ||
Db 49198 AAGGAGCTTTGTAACCTGTTGGAGAGGATTTATCTTCCGATTCACATGATGATG 49257
QY      63 gtagatccggtgcatagcgttgcctaaagggagagatccgttgcctcatatga 122
        || || || || || || || || || || || || || || || || || || || ||
Db 49258 GTTAGCTCTAGTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 49317
QY      123 gatgaattgatcccgagagatcatcactgctatagaatgagatccattcagaa 182
        || || || || || || || || || || || || || || || || || || || ||
Db 49318 GAGGAGCTGATTCCTACTCTGGACAAATCACAGGCCGCCGATGTGATGATGATGATG 49377
QY      183 ctgaataaagccactagaagaatcatcacttgccttgccttgccttgccttgcct 242
        || || || || || || || || || || || || || || || || || || || ||
Db 49378 ATGAATGCTGTACACAGAAACACATTTTCCCTTACCATTTATGATGATGATGATG 49437
QY      243 aggttgcctaaatcacacacttgccttgccttgccttgccttgccttgccttgc 302
        || || || || || || || || || || || || || || || || || || || ||
Db 49438 ATGTTAGCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATC 49497
QY      303 ccaattgcacaatctgatcagagagaacacacttgccttgccttgccttgccttgcct 362
        || || || || || || || || || || || || || || || || || || || ||
Db 49498 CCGATTCATTCGAGATGACACAGAGAAACGACTTTCACCTGCTCTTATGATGATGATGCT 49557
QY      363 tatagagtagtgccttgccttgccttgccttgccttgccttgccttgccttgccttgc 422
        || || || || || || || || || || || || || || || || || || || ||
Db 49558 TATCGGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 49617
QY      423 gctatctctcaactcttgccttgccttgccttgccttgccttgccttgccttgccttgc 482
        || || || || || || || || || || || || || || || || || || || ||
Db 49618 TCTATCTTCAACAGACATGATGATGATGATGATGATGATGATGATGATGATGATG 49677
QY      483 tagaggtctcttgccttgccttgccttgccttgccttgccttgccttgccttgccttgc 542
        || || || || || || || || || || || || || || || || || || || ||
Db 49678 TATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 49737
QY      543 gacacacatctgcttgccttgccttgccttgccttgccttgccttgccttgccttgc 602
        || || || || || || || || || || || || || || || || || || || ||
Db 49738 GAGAAACATCTGATTTTAAATTTGGGAAATGTCACCTTATGATGATGATGATGATG 49797
QY      603 ttggagacataaattctcgaagaaggtatgaagtcgaataaggaaggtatgaatc 662
        || || || || || || || || || || || || || || || || || || || ||
Db 49798 CTCGAGACAGAGATTTCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 49857
QY      663 gacaacatgcctacccacagatcaagaagtataagaagttcccttgcataagtg 722
        || || || || || || || || || || || || || || || || || || || ||
Db 49858 ACAAGTCTTCAGAGCGCTGATATGATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA 49917
QY      723 tctcatagaaggttatcaagaattca 750
        || || || || || || || || || || || || || || || || || || || ||
Db 49918 TTCTACAGGAGATTCATCAAAAGCTTCA 49945

```

RESULT 7
 LOCUS AC007187 95685 bp DNA
 DEFINITION Arabidopsis thaliana chromosome II section 63 of 255 of the
 complete sequence. Sequence from clones T1806, F7E22, F23M2.
 ACCESSION AC007187 AE002093
 VERSION AC007187.4 GI:6598701
 KEYWORDS
 SOURCE
 ORGANISM

thale cress.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 95685)
 Lin X., Kaul S., Kounsley S.D., Shea T.P., Benito M.-I., Tom M.C.D.,
 Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E.,
 Feldlyum T.V., Buell C.R., Ketchum K.A., Lee J.J., Renning C.M.,
 Koo H., Moffat K.S., Cronin L.A., Shen M., Vanden S.E., Umayam L.,
 Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
 Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,

Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and
 Venter, J.C.
 Sequence and analysis of chromosome 2 of the plant *Arabidopsis*
thaliana
 JOURNAL Nature 402 (6763), 761-768 (1999)
 MEDLINE 20083487
 PUBMED 10617197
 REFERENCE 2 (bases 1 to 95685)
 LInX.
 Direct Submission
 Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 On Dec 17, 1999 this sequence version replaced gi:4733998.
 The sequence and annotation of chromosome 2 were merged from those
 of the individual clones on this chromosome after removing
 overlaps. For detailed information, please see the TIGR web site
 (<http://www.tigr.org/tcdb/at/at.html>).

Genes were identified by a combination of three methods: Gene
 prediction programs including GRAL.
 (<ftp://arthur.epm.ornl.gov/pub/xgrail>), Genefinder (Phil Green,
 University of Washington), Genscan (Chris Burge,
<http://genomic.stanford.edu/GENSCAN.html>), and NetPlantGene
 (<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the
 complete sequence against a peptide database and plant EST
 databases at TIGR, and manual curation based on those analyses.
 Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as 'unknown' proteins.
 Genes without protein or EST similarity, that are predicted by two
 or more gene prediction programs over most of their length are
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are
 predicted by tRNAscan-SE (Sean Eddy,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were
 identified by RepeatMasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are
 numbered from the top to bottom of the chromosome.

We thank the CSHL/Mashu/ABI consortium for sequencing BAC clones
 F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone
 F13P4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
 and Satoshi Tabata for helpful assistance. In addition, we would
 like to thank the TIGR Bioinformatics Department, especially Lixin
 Zhou, Hanif Khailak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
 Peterson, Michael Holmes, and Delwood Richardson for software and
 database support.

This work was supported by the National Science Foundation,
 Department of Energy and the US Department of Agriculture.

Address all correspondence to: atetigr.org.

FEATURES
 Source

Location/Qualifiers
 1..95685
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="II"

misc_feature
 /note="Sequence from clone T1806"
 complement(302..95484)
 /note="F7E22.16: TNP2 protein of the Tam1 family of
 transposons [Antirrhinum majus]"
 344..2412
 /gene="At2g11670"

gene
 /pseudo
 373..434
 /rpt_family="POLY_A"

repeat_region
 3213..3703
 /note="DNA sequence related to Athila retrotransposon"
 <3898..>4803
 /gene="At2g11680"

gene

<3898..>4803
 /gene="At2g11680"
 /note="F7E22.15"
 3898..4803
 /gene="At2g11680"

CDS

/codon_start=1
 /product="putative retroelement pol polyprotein"
 /protein_id="A028680.1"
 /db_xref="GI:4734002"

mRNA

/translation="MVRGIVLGHKISKGGVGNKAKVDVMMLOLPKTKDIRSFLG
 HAGRRFRIDEKSLKAPLRLRLCKRAEFTFECECTAFKSTKALVSAPIVOAPRD
 YPEELMCDSDVAVGAVLGGKIDRKLVHITYASGTDPTVDRATTEKELLAVPAFE
 KFSYLVGSVTVYTDHALRLHYAKDKTPRLRLNLILOEDMEIVKKGLENCA
 DLSRMRIEDEVPIDDSMPEEQMLAIQOLNESAOINKSLDOVCAYEDKLMYADHVKY
 LFGSEPPNLSYEKKFFKDIHNF"
 complement(join<7068..7358,7445..7738,7824..8150,
 8234..>8380))
 /gene="At2g11690"

gene

/complement(<7068..>8380)
 /gene="At2g11690"
 /note="F7E22.14: related to En/Spm transposon family of
 maize"
 complement(join(7068..7358,7445..7738,7824..8150,
 8234..>8380))
 /gene="At2g11690"

CDS

/codon_start=1
 /product="En/Spm-like transposon protein"
 /protein_id="A028679.1"
 /db_xref="GI:4734001"

misc_feature

/translaction="WGRGVATRKRAFLQARDAYOKLEAOALILNRINDLOEVCGL
 AKGKSDVSISETSKVATITFCGLDWCNSDILVGESEFCSAERTYKIGTIPLKNAAY
 VIVNEKEDAPVWRPVATITFSLNMGKTIPIWSPNKITVDNANVNSPTENHTAGSSNP
 PRLILFDVNWEDILVAGHLISDPEKELANNIPICPNALVAVKIKFIDKAYLWPR
 TSEMILMGDNLHELIPWIDRIQISGNSPTESSPSPASCSMAKTTKSDQKRLIL
 LDGSGSEIYAEGRICSTNPBDKHFPLPDPASKVWVYLKIGSAKVMRPNSEVHF
 GDALGTVAMPNDRLRPM"
 complement(810..14100)
 /note="putative retroelement. Structure: LTRa,
 retroelement, LTRB"

gene

9615..11655
 /gene="At2g11700"
 /note="F7E22.13"
 /pseudo
 complement(11014..11053)
 /rpt_family="(CAAAA)n"

repeat_region

12552..12623
 /rpt_family="POLY_A"

repeat_region

14100..14907
 /rpt_family="POLY_A"

repeat_region

/note="DNA sequence similar to Athila retrotransposon"
 complement(15034..16032)
 /gene="At2g11710"

gene

/note="F7E22.18"
 /pseudo
 complement(16701..19304)
 /gene="At2g11720"

gene

/note="F7E22.17: TNP2 protein of the Tam1 family of
 transposons [Antirrhinum majus]"
 /pseudo
 19615..21837
 /gene="At2g11730"

gene

/note="F7E22.12"
 /pseudo
 complement(22231..23920)
 /gene="At2g11740"
 /note="F7E22.11: TNP2 protein of the Tam1 family of
 transposons [Antirrhinum majus]"

gene

/pseudo
 complement(25972..31340)
 /note="putative retroelement. Structure: LTRa,
 retroelement, LTRB"

misc_feature

26459..29209
 /gene="At2g11750"

gene

/note="F7E22.10"

Query Match	50.3%	Score 380.8;	DB 8;	Length 95685;
Best Local Similarity	69.1%;	Pred. No. 1.6e-88;		
Matches 520;	Conservative 0;	Mismatches 232;	Indels 0;	Caps 0

Query Match	50.3%;	Score 380.8;	DB 8;	Length 95685;
Best Local Similarity	69.1%;	Pred. No. 1.6e-88;		
Matches 520;	Conservative 0;	Mismatches 232;	Indels 0;	Caps 0

	RESULT	8						17-JAN-2003
AX059456/c								
LOCUS	AX059456		40480 bp	DNA	PAT			
DEFINITION	Sequence	189	from Patent W0005535.					
ACCESSION	AX059456							
VERSION	AX059456.1		GI:12311561					

BASE COUNT	ORIGIN	12306 a	7912 c	8765 g	11297 t	200 others
/organism="Arabidopsis thaliana"						
/db_xref="taxon:3702"						

Query Match	50.2%;	Score 380;	DB 6;	Length 40480;
Best Local Similarity	68.9%;	Pred. No. 2.4e-88;		
Matches 521;	Conservative	0;	Mismatches 235;	Indels 0;
				Gaps 0

QY	2	aaagaggtctgtgaagctccctcgatbaaggtattactatcaatggtgcacatagcatg	61
Dp	32944	AAAAGAAATTTTGAACCTCCTTGAGGCTGCTGTCATCTTCTATCTTGATGATGCTCTTG	32885
QY	62	gttgaatccggtgcatgaacgcttccctaaagaaggacattaccgtgtgtccctatgataa	121
Dp	32884	GGTTTCTCCAGTGCATTTGGCTCCCTAAAAAGGATGGAATGACTGTTGTCAAAATGAAA	32822
QY	122	gaatgaatgtcccccagagatatactccctgcatagatggtatcgattccatgaaga	181
Dp	32824	ACATGACTAAATCCCTACTAGACATATACTGGTCATAAAGTGCATTTGATTATAGGA	32765
QY	182	actgaataaagccataggaagaatcatcaaccttgcctttatccagacattagtga	241
Dp	32764	GTTAATGGTCATCTAGSAAAGATCATTTTCCTTACCATTCAATTGACCAATAGCTTGA	32705
QY	242	aaggtgtccaaactcaacacactctgtcttctctagaaggttaattctagtttccccaat	301
Dp	32704	ACGTTTACTATATCAATCATATACTATTGCTTTCTTGATGATATATAGTGTTCTTCAAT	32645
QY	302	accagltgcacatctgtctgagagaanaaacatttcaaccgaccttcgtaaccttgc	361
Dp	32644	ACCAATTAACCTTAATGATCAAGAAAAACCATTTTCACATGCTCCCTACGGAACTTTTGC	32585

QY 362 ttatagacgtatgcttcttgctatgtaatgacacgtcccttcaagaatgatatg 421
 Db 32584 TTATAGAGAAATGCTATGGTTTATGTAATGCTCTCCACCATTTTCAGAGGTATGAC 32525
 QY 422 ggcctatcttcaactcttgtagaaatgtagtgcgtttcagtgagttccttccgt 481
 Db 32524 CTCTATATTTTTCAGATTTTAATGAGAGAGTGTGAGGTTTTCATGACGATTTTTCGCT 32465
 QY 482 ttaaggctctcttcttgatgtagtcccaagcaactgacgtcttcaagagatgtaa 541
 Db 32464 CTAATGGCCCCCTCTTCTCTCTCATGTTTGTGTAATCTTGCAAGGTATGACAGGTGCA 32405
 QY 542 agacaccaaactctctctgtaatgtaggagagatgacattatgtagtaagagatcgt 601
 Db 32404 ACAGACGAAATCTTGTCTCAATTTGGGAAAGTGTATTTTCATGAGGAGGATGAT 32345
 QY 602 cttaaggacataaattcttgaaagaggtatgtaagtcgataaggttaaggttgatgcaat 661
 Db 32344 GTTGGCTCACAGATATTCAGAGAGGATTAAGAGTTGCAAGAAATTTGAAGTATGAT 32285
 QY 662 cgacaaatgacatacccccacagatatacaaggtatcaaggtcttctgtagcatggtg 721
 Db 32284 GATGAGATTGCAAGCCCAAAAACGGTGAAGACATCAAGATTCCTGTCATGCTGG 32225
 QY 722 ttctatagaaggttatcaagaattcccaaggt 757
 Db 32224 GTTCTACAGAGATTTTAAAGACTTTCACAGAT 32189

RESULT 9

AC016828 76911 bp DNA HTG 29-SEP-2000
 LOCUS Arabidopsis thaliana chromosome 3 clone IGF-F21A14, *** SEQUENCING
 DEFINITION IN PROGRESS ***, 1 ordered pieces.
 AC016828
 AC016828.5 GI:10337645
 VERSION HTG: HTGS_PHASE2.
 KEYWORDS thale cress.
 SOURCE Arabidopsis thaliana
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

Lin, X., Kaul, S., Town, C. D., Benito, M. I., Creasy, T. H.,
 Romling, C. M., Koo, H., Fujii, C. Y., Utterback, L. J., Barnstead, M. E.,
 Bowman, C. L., White, O., Nierman, W. C. and Fraser, C. M.
 Arabidopsis thaliana 'IGF' BAC 'F21A14' genomic sequence near
 marker '1'

TITLE

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 76911)
 AUTHORS Lin, X. and Kaul, S.
 TITLE Direct Submission

JOURNAL

COMMENT

Submitted (08-DEC-1999) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA, xlin@igf.org
 On Sep 29, 2000 this sequence version replaced gi:6692721.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however, the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 76911: contig of 76911 bp in length.
 * Location/Qualifiers

FEATURES

source

1. 76911
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="IGF-F21A14"
 BASE COUNT 21697 a 16043 c 16139 g 22185 t 847 others

ORIGIN

Query Match 50.2%; Score 380; DB 2; Length 76911;
 Best Local Similarity 68.9%; Pred. No. 2, 6e-88;
 Matches 521; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

QY 2 aaagaaggtgtagagcttcctgtagaaggtatcatcatcagatgtagcatgcatg 61
 Db 16644 AAAAGAAATTTTGAATCTCTTGAAGCTGTGTCATCTATCTCATCTGATAGTCTTT 16703
 QY 62 ggtagtcggtgtagatagcgttccaaagagagagagcatcaccgttgcctaaatgataa 121
 Db 16704 GGTTCCTCCAGTGCATTTGCGTCCCTAAAGGGGGAATGACATGTTGTCAAAAATGAGAA 16763
 QY 122 gtagtaattgtagcccgagagagattcaactgtagtataagatgtagtgcattcaggaa 181
 Db 16764 AGATGAACATTAATCCCTCTCTAGAACCTTACCTGTCATGAGATGTGCAATTAATAGGA 16823
 QY 182 actgaataaagccactaggaagaatcatcattacccttgcctttaaagacatatgtaga 241
 Db 16824 GTTAATCTCTCATCTAGAGAAAGATCATTTTCTTTACCATTCATTCACCAATAGCTTGA 16883
 QY 242 aaggttgcacaactcaacacacactctgcttctagacggtatctagtcttcccaat 301
 Db 16884 ACGTTTACCTAATCATCATCAATCAATGCTTCTTGATGAGATATAGTGCTTTCTTCAAT 16943
 QY 302 accagttgcacaactgtagtagagaaacacacttcaacccttgccttgcattcctgtag 361
 Db 16944 ACCAATTCACCCCTAATGATCAAGAAAAACCATTTTACATGATGCTTACAGAACTTTTGC 17003
 QY 362 ttatagacgtatgcttcttgctatgtaatgacacgtccacacttcaagaatgatatg 421
 Db 17004 TTATAGAGAAATGCTATGGTTTATGTAATGCTCTCCACCATTTTCAGAGGTATGAC 17063
 QY 422 ggcctatcttcaactcttgtagaaatattgtagaggttcaatgtagtgcattcctcgt 481
 Db 17064 CTCTATATTTTTCAGATTTTAATGAGAGAGATGCTGAGGTTTTCATGACGATTTTTCGCT 17123
 QY 482 ttaaggctctcttcttgatgtagtgcacgaacttgcagacttgcagatgtagatgtaa 541
 Db 17124 CTAATGGCCCCCTCTTCTCTCATGTTTGTGTAATCTTGACAGGTTGACACAGTGGGA 17183
 QY 542 agacaccaaactctctctgtaatgtaggagaggtgacacttattgtagtaagagcatcgt 601
 Db 17184 AAGAGAGAAATCTTGTTCATATGGGAAAGGTGATTTTCATGAGGAGAAAGCATAGT 17243
 QY 602 cttaaggacataaattcttgaaagaggtatgtagatgtagtaaggtatgtagcat 661
 Db 17244 GTTGGGTCAAGATATCAAGAAAGGATATAGGTTGACAAAGGAAATTTGAAGTAT 17303
 QY 662 cgacaaatgacatacccccacagatatacaaggtatgaaggttctctgtagtagtgg 721
 Db 17304 GATGCAATTGCAAGCCCAAAAACGGTGAAGACATCAAGATTTCTTGGTCAATGCTGG 17363
 QY 722 ttctatagaaggttatcaagaattcccaaggt 757
 Db 17364 GTTCTACAGAGATTTTAAAGACTTTCACAGAT 17399

RESULT 10

AB046431/c 85962 bp DNA PLN 18-APR-2001
 LOCUS Arabidopsis thaliana DNA, chromosome 5 centromere region,
 DEFINITION clone: T3B16.

ACCESSION AB046431.1 - GI:9971602
 VERSION AB046431.1 - GI:9971602

KEYWORDS

SOURCE Arabidopsis thaliana (cultivar: Columbia) DNA, clone: T3B16.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (sites)
 AUTHORS KumeKawa,N., Hosouchi,T., Tsuruoka,H. and Kotani,H.
 TITLE The size and sequence organization of the centromeric region of Arabidopsis thaliana chromosome 5
 JOURNAL DNA Res 7 (6), 315-321 (2000)
 MEDLINE 21082928
 REFERENCE 2 (bases 1 to 85962)
 AUTHORS Kotani,H. and KumeKawa,N.
 TITLE Direct Submission
 JOURNAL Submitted (21-JUL-2000) Hirokazu Kotani, Kazusa DNA Research Institute, Lab. Chromosome Research, I-1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:kotani@kazusa.or.jp, Tel:81-438-52-3920, Fax:81-438-52-3921)
 FEATURES
 source 1. 85962
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /map="centromere"
 /clone="T3B16"
 1. 2000
 /note="overlap to BAC F06024, please refer for analysis and annotation"
 2000. 11119
 /note="similarity to 180 bp centromeric repeat"
 11120. 13109
 /note="similarity to 180 bp centromeric repeat"
 13110. 14148
 /note="similarity to 180 bp centromeric repeat"
 14149. 14678
 /note="similarity to 5S rDNA (1 copy) and spacer (1 copy)"
 14679. 16228
 /note="similarity to fragments of Athila retrotransposons"
 16229. 17527
 /note="similarity to fragments of Athila retrotransposons"
 17528. 18305
 /note="uncharacterized tandem repeat"
 18306. 18350
 /note="similarity to fragments of Athila retrotransposons"
 18361. 19273
 /note="similarity to fragments of Athila retrotransposons"
 19274. 20452
 /note="similarity to fragments of Athila retrotransposons"
 20453. 28235
 /note="similarity to fragments of Athila retrotransposons"
 28236. 30040
 /note="similarity to fragments of Athila retrotransposons"
 30041. 31528
 /note="similarity to fragments of Athila retrotransposons"
 31529. 33152
 /note="similarity to fragments of Athila retrotransposons"
 33153. 44423
 /note="similarity to fragments of Athila retrotransposons"
 44424. 45933
 /note="uncharacterized tandem repeat (2 copies)"
 45934. 47691
 /note="similarity to fragments of Athila retrotransposons"
 47692. 55569
 /note="similarity to fragments of Athila retrotransposons"
 55570. 57327
 /note="similarity to fragments of Athila retrotransposons"
 57328. 61906
 /note="similarity to fragments of Athila retrotransposons"
 61907. 63599
 /note="similarity to 5S rDNA (2 copies) and spacer (2 copies)"
 63600. 64534
 /note="similarity to 180 bp centromeric repeat"
 64712. 68805
 /note="similarity to 180 bp centromeric repeat"
 68806. 70552
 /note="similarity to fragments of Athila retrotransposons"
 70553. 81073
 /note="similarity to fragments of Athila retrotransposons"
 81074. 82821
 /note="similarity to 180 bp centromeric repeat"
 82821. 85962
 /note="similarity to 180 bp centromeric repeat"
 BASE COUNT 25983 a 16785 c 17774 g 25420 t
 ORIGIN
 Query Match 50.2%; Score 380; DB 8; Length 85962;

Best Local Similarity 68.9%; Pred. No. 2,66-88;
 Matches 521; Conservative 0; Mismatches 235; Indels 0; Gaps 0;
 QY 2 aaaaaggttctgaagctccctgcatgaagatattatctatcattgctcatagcatg 61
 Db 77876 AAAAGAAATTTGAAACAGTGTGATGCTGTGTATCACTACCTACCTGATGACTTG 77817
 QY 62 ggtagtcggtgcatgacgtctctctaaagaaggagcatccggttccctaatgata 121
 Db 77816 GGTTCCTCAGTGTGATGCTGTGCTTAAAGGAGGATGATGCTGTGCTTAAAGGAA 77757
 QY 122 ggaatgaattgctcccgcaagagatattacatgctgctgcatagatgctgcatgaa 181
 Db 77756 AGATGAACGTGATACCTACTACACTATACCTGCTGCTATGATGCTATGATGAA 77697
 QY 182 actgaataaagccatggaagaagatcattacccttgctcttatacgacatagctaga 241
 Db 77696 GTTGAATGCTGCATCTGAGAAAGATCATTTCTTTACCATTCATGACCAATGCTTGA 77637
 QY 242 aaggtgtccaaactcaacacacgtctctctctagaaggtatctagttcccaat 301
 Db 77636 ACGATTAGCTATCATCATATTTATTCCTTTCTTGATGATGATGATGATGATGATGAT 77577
 QY 302 accagttgcaaatctgacaggaagaacacacacacacacacacacacacacacacacac 361
 Db 77576 ACCAATTCACCCCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 77517
 QY 362 ttatagaacgtatgcttggtgcttatagtatgaacacacacacacacacacacacac 421
 Db 77516 CTATAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 77457
 QY 422 ggcatactcttaac 481
 Db 77456 CTCTATATTTTCAGACTTAATTGAGGATGATGATGATGATGATGATGATGATGATG 77397
 QY 482 ttacggctctcttctgattgattgctcctcaacacacacacacacacacacacacacac 541
 Db 77396 CTATGCGCCCTCTTTCCTTCATGTTGTTGATGTTGTTGATGTTGTTGATGTTGATG 77337
 QY 542 agac 601
 Db 77336 GGAAGCAGAACTTGTCTCAATTTGGAAGATGATGATGATGATGATGATGATGATG 77277
 QY 602 ctaagacataaaattctgaagaagatgataagacacacacacacacacacacacacacac 661
 Db 77276 ATTGGGTCACAAATATTCAGAGAAAGTATGAGGTTGACAAAGAAAGTATGAGTAAT 77217
 QY 662 cgacaaaatgcataccacacagatatcaaaagtataagaagttctcttgatgctg 721
 Db 77216 GATGACGCTGCACGCACCAAAACGGAAGGACATCAGAACTCTTGCTATCTGTG 77157
 QY 722 ttctatagaagttatcaaaagattcacaaggt 757
 Db 77156 ATTCTATAGAAATCATTAAGACTTCTCCAGAT 77121
 RESULT 11
 AB046426/c
 LOCUS AB046426
 DEFINITION Arabidopsis thaliana DNA, chromosome 5 centromere region,
 clone:F28N5.
 ACCESSION AB046426
 VERSION AB046426.1 GI:9971597
 KEYWORDS centromere.
 SOURCE Arabidopsis thaliana (cultivar:Columbia) DNA, clone:F28N5.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (sites)
 AUTHORS KumeKawa,N., Hosouchi,T., Tsuruoka,H. and Kotani,H.
 TITLE The size and sequence organization of the centromeric region of Arabidopsis thaliana chromosome 5

JOURNAL
 DIRECT SUBMISSION
 Submitted (12-JAN-2000) Arabidopsis thaliana genome center,
 Department of Biology, University of Pennsylvania, 38th and
 Hamilton Walk, Philadelphia, PA 19104-6018, USA
 5 (bases 1 to 118986)


```

/db_xref="taxon:3702"
/clone="T6L9"
/chromosome="1V"
/map="unknown"
1206..2250
/gene="T6L9.5"
join(1206..1364,1437..1532,1614..1776,2072..2250)
/gene="T6L9.5"
/codon_start=1
/evidence=not_experimental
/protein_id="A038217.1"
/db_xref="GI:5032269"
/translation="MANEVSILCPDLGIPLYGSDPPTSYLPHYVYEPKPK
KESYDFADKYLEKRRACDERHILGLRQLEAKIAQMERKATLGSGRPDPNRVY
KVPFDESGDEEDINTMKRVKTNPMKAPRRLQNMEDLYSTETVDSADGMI
REVPKYPDORPPLRGYFRSEIWTPTCTLIKDLPERDQ"
complement(3014..3990)
/feature="pseudogene; contains similarity to retroviral
polyproteins"
1394..15837
/gene="T6L9.4"
join(1394..15454,15589..15837)
/gene="T6L9.4"
/feature="similar to Arabidopsis thaliana putative En/Spm
transposon protein (GB:AC005396)"
/codon_start=1
/evidence=not_experimental
/protein_id="A038214.1"
/db_xref="GI:5032266"
/translation="MDAYKKRCMDWHHGESNSRAESPTKVYQNNDELYOAEFLDE
EPAYVGEIETAEEDDKDEDEFLAKIADEEPLPSCNYSKLSAIVSLFELNKGW
SDKSFNDLELEPEMLPKDNLVHTSLYKVKFKLSFDMGSKIKIACVNDCLFPRNK
KUDSPCKNASRKMTNLHIGELKGVOKVLRTPITRGLGMRKMDLKHMS
LMSDEKLCHPDLVTWQNNAKYPSEGEERNLRHGLSTGDFPNNKRTISCWPD
LCKKENIMTLILIPQOQNNIDVLEPIEDICHLMDGELTYAASSTEDLKA
MLMTISDPFPAVGLAGCKVGMCCPLGKNTDSMLKFSRKHVYMCHILGLPTTH
FRMKSMFSGKAENGRKRRLISGOETYNLKNKNNFETLKOSASKRRTVINIADSD
SDOSSESEBEDQDVDEELSRMKKSLFEPKLPWQELLVRHNDVHVRVNAASI
ISTLHKWKSQDGLNARKDLEVLIRKDHRTKGRITLPAALMSLEICARMS
1K"
1677..18753
/feature="pseudogene"
complement(22519..27395)
/feature="contains similarity to retrotransposon-like
proteins"
29824..33456
/gene="T6L9.2"
join(29824..29837,29873..30092,30287..30343,30437..30687,
30777..30965,31095..31230,31477..31584,31668..31913,
32000..32038,32118..32417,32580..33119,33205..33456)
/gene="T6L9.2"
/feature="contains similarity to Arabidopsis thaliana
hypothetical protein (GB:AC005396); maybe a pseudogene"
/codon_start=1
/evidence=not_experimental
/protein_id="A038216.1"
/db_xref="GI:5032268"
/translation="MYLLEPYILVSOAKOVYLRDESSPPYVALRGSSRPSKEE
LDSVAIGLPNVDIDADLEDERENRTDCEIYNAAKANEVEITIGIEVONDED
HLEPDVAMPDQVDDGLKVNNEEPHGLVISEGDCVEPRKRLRRGTRLDIAKD
EOKYRARIADDSOLSVLIIMVNPQVEFTNMPGLDGLSVGLAVGRVHP
VTEERTKIGELRTYLVMSVOYSFVMSROYTNREKATNOERMMNPKNISFED
RRPVKRTSQEFKRTASADSEFVKRLKVWVKSRTKDAPVNTNATEKIKKKEIYVS
DNPSNEADDSLSQLGPNRGLRAMGNNMKRTACGONHMGCAEMOENLOK
KVNLEIVLEKMNQROSAEVBENSARSVNKTQPKCFLLOHSGSDAYAAAGCTISS
DDELVNDRLGPTDVKFLIDAIEPEALMRAAMFLITEAVGIMLWPAAMCSL
EEDLOEDIAICQEBNALVNGLPGLHVAVKLFVDVVLDPPEFTIMPRTNVYLEBCD
MSFVAMPVNVKVIENPADATYOKSPLOTSAARGLSEATTSYGKSAASVSGPSGL
KSARSKKAVASGAASGOTAGYKSEIPEAKGTISAVATYSKIAEKSIAGSKS
PVKSQLDSQLHRLRSRKGVAEVIKTHKCKIMDISRKQVAVAGRVHSDPDQK
QVRLGDNARVAVDVKIDDATVWRASDEIETMRDSSLSSTA"
50747..51653
/gene="T6L9.1"

```

```

CDS
join(50747..50820,51398..51426,51475..51653)
/gene="T6L9.1"
/codon_start=1
/evidence=not_experimental
/protein_id="A038215.1"
/db_xref="GI:5032267"
/translation="MOTFEIVIGYVLLGIPISIRNENMFELCWVYKEDDEEEN
DEDEKEKEEYKNYADGESNDNGNENMLTIKIMRLNMKIIORHVLVMI"
BASE COUNT 16547 a 10322 c 9855 g 17849 t
ORIGIN

Query Match 50.0%; Score 378.4; DB 8; Length 54573;
Best Local Similarity 68.8%; Pred. No. 6,4e-88;
Matches 520; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

QY 2 aaaaaggttgtagaagctccgtgtagaaggtatataatcatatgctatgcatgcatg 61
Db 25180 AAAAGAAATTTTGAACATGCTGATGCTGGTGCATCTACCTCATCTGATGTA 25121

QY 62 ggtgagtcggtgcatagcgtctcctaagaaggaggagattacggtgctccatgtaa 121
Db 25120 GGTTCCTCCAGTCATGTCGTCCTTAAGAGGAGGATGACGTGTCTCAAAAATGAAA 25061

QY 122 ggtgaaatgtagcccgagagattatcatcgtgcatagagtgtagatcgaattcagaa 181
Db 25060 AGATGAATGATCCCTACTACTAGAACGTCATAGATGATGATGATTAAGAA 25001

QY 182 actgaataaagccacatggaagaatcatcacccttgcccttatacgacatattgcaga 241
Db 25000 GTTGAACTGTCACCTGAGAAAGATCATTTTCCTTACCATTCATGACCAATGCTTGA 24941

QY 242 aaggtgtccaaactcaacacactctccttctagaagttatctatgcttctcccaat 301
Db 24940 ACGTTTGCTATATCATCATATTTATGCTTCTTGATGAGATAGAGTGTTCTTAAAT 24881

QY 302 accaggtgcacaatctgtagcagagaanaaccacttcaactgccttccttcggtacattgc 361
Db 24880 ACCAATTCACCTATATGATGATCAAGAAAACACATTCACGTGCTTATGGAACCTTTGC 24821

QY 362 ttatagaagtagcttcttgcttataatgaatgcacgtgcacacttcaagaagtagat 421
Db 24820 TTATAGAGAAAGCCATTTGTTATGCAATGCTCTGCAACATTTCAAGAGTATAGC 24761

QY 422 ggcatacttcctaaccttgtagaanaatctgcagagtttcttctatgtagacttccgt 481
Db 24760 CTTATATTTTATAGCTTAATGAGAGAGATGTAAGAGTTTTCATGAGCATTTGTCGT 24701

QY 482 ttacggtctctcttgtagatgtagcctcaagaacactgtagatcgtatcagaagatgaa 541
Db 24700 CTATGGCCCTCTCTCTCTCATGTTGTGAAATGTTGGCAGGGATTTGACCAAGTGCGA 24641

QY 542 aagaccacatctgctctgtagaagtgaggagagagcactttagttaaagaagcatcgt 601
Db 24640 AAGAGACGAATCTTGTTCTCAATTTGGAAAGTGTCATTTTCATGCGGAAGAGCATGT 24581

QY 602 cctagaacataaattctcgaagaagtagttagtcogataagcctaaggttagatgcaat 661
Db 24580 GTTGGGTGCAAGATATTAAGAGAGGATATAGAGGTGCAAAAGGAAAGTTGAAGTAT 24521

QY 662 cgacaaaatgcataccacacagataatcaagaagtagaagttccttgctgtagtgcg 721
Db 24520 GATGCGAGTTGCGCCCAAAAACGGTGAAGGAATTCAGAAATTCCTTGTCATGCTGG 24461

QY 722 ttctatagaaggttatcatcaagaattcacaaggt 757
Db 24460 GTTCTATAGAGTTTATTAAGACTTCTTAAGAT 24425

RESULT 15
LOCUS AB026642/c
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, Pl clone: MBD5.

```

```

ACCESSION      AB026642.1
VERSION        AB026642.1
KEYWORDS       GI:4757398
SOURCE         Arabidopsis thaliana (strain:Columbia) DNA, clone_11b:Mitsui P1
               clone:MED5.
ORGANISM       Arabidopsis thaliana
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
               Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi
               1 (sites)
REFERENCE      Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.
AUTHORS        Structural analysis of Arabidopsis thaliana chromosome 3. I.
TITLE          Sequence features of the regions of 4,504,864 bp covered by sixty
               p1 and TAC clones
JOURNAL        DNA Res. 7 (2), 131-135 (2000)
MEDLINE        20277480
REFERENCE      2 (bases 1 to 77287)
AUTHORS        Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.
TITLE          Direct Submission
JOURNAL        Submitted (28-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
               Institute, Department of Plant Gene Research; 1532-3, Yana,
               Kisarazu, Chiba 292-0812, Japan (E-mail:yakamukazusa.or.jp,
               Tel:81-438-52-3935, Fax:81-438-52-3934)
               Address for correspondence: kaoskazusa.or.jp
               For the latest information on annotation of this clone, please see
               http://www.kazusa.or.jp/kaos/cgi-bin/agd-graph.cgi?c=MED5
               Genes with similarity to proteins in the databases are described in
               'product' or 'note' qualifiers. Genes that have no significant
               protein similarity are described as 'unknown protein'
               The software programs used to predict genes include: Graal
               (Informatics Group, Oak Ridge National Laboratory,
               http://compbio.ornl.gov/Graal-1.3/),
               GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
               NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
               Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
               SplicePredictor (Volker Brendel, Stanford University,
               http://gremmlin.zool.iastate.edu/cgi-bin/sp.cgi).
               Genes encoding tRNAs are predicted by tRNAscan-SE
               (Sean Eddy, Washington University School of Medicine, St. Louis,
               http://genome.wustl.edu/eddy/tRNAscan-SE/).
               This sequence may not be the entire insert of this clone. It may be
               shorter because we remove overlaps between neighboring submissions.
               The 5' clone is MED16 and the 3' clone is F21A17.
FEATURES       source
               1..77287
               /organism="Arabidopsis thaliana"
               /strain="Columbia"
               /db_xref="taxon:3702"
               /chromosome="3"
               /clone="MED5"
               /clone_11b="Mitsui P1"
               join(689..829,1056..1214)
               /note="gene_id:MED5.1
               unknown protein"
               /codon_start=1
               /evidence=not_experimental
               /protein_id="BAB01785.1"
               /db_xref="GI:9293882"
               /translation="MCMAYLVFRYGCWSSNRGKTEHEVAMPNDVFSKNTLSQPRS
               HARGRWTIOSRIHCHSIDRKTEPLKSSLIINDLHWEAFHSPSDANDJFGRIM"
               complement(join(2226..2351,2434..2661,2796..3083,
               3159..3326))
               /note="gene_id:MED5.2"
               /codon_start=1
               /evidence=not_experimental
               /product="transposon protein-like"
               /protein_id="BAB01786.1"
               /db_xref="GI:9293883"
               /translation="MFTSFFVACVWRFRDGRSLVKRTIRIFLOKDPAPYNNMNCVP
               VDKRERLFLPAKTHHMDPLLTGTVOYTFNFIKRLDWDVSTARTTREGPPWIGETL
               WGTGCAFWIDTAOKRSRTYSKARLSQNGIGPVHVSQSPFOIDELKEKIGRPV
               HLGSEVFLTHTRSDGSEFVDSKSEKIAQAYQONVDRSLALASASVSDGSSRPPELT
               LDVYTAIFLESTEDKSRGNPYGLCKLDTLGSANRNHSGSSSSFOALEER"
               CDS
               complement(join(4457..4848,4900..5115))
               /note="gb|AAD28667.1
               gene_id:MED5.3
               similar to unknown protein"
               /codon_start=1
               /pseudo
               /evidence=not_experimental
               complement(6870..15564)
               /note="gene_id:MED5.4"
               /codon_start=1
               /pseudo
               /evidence=not_experimental
               /product="Athila retroelement ORF1 protein-like"
               complement(join(16691..17082,17134..17349))
               /note="gb|AAD28667.1
               gene_id:MED5.5
               similar to unknown protein"
               /codon_start=1
               /pseudo
               /evidence=not_experimental
               complement(18583..18607,18971..19069))
               /note="gene_id:MED5.6"
               /codon_start=1
               /pseudo
               /evidence=not_experimental
               /product="transposon protein-like"
               complement(19942..23841)
               /note="gene_id:MED5.7"
               /codon_start=1
               /pseudo
               /evidence=not_experimental
               /product="transposon protein-like"
               complement(1942..23841)
               /note="gene_id:MED5.8"
               /codon_start=1
               /pseudo
               /evidence=not_experimental
               /product="transposon protein-like"
               complement(26389..27150,27293..27596,27715..28429,
               28770..28877,29127..29173,29504..29571))
               /note="gene_id:MED5.9"
               /codon_start=1
               /evidence=not_experimental
               /product="Ac-like transposase"
               /protein_id="BAB01787.1"
               /db_xref="GI:9293884"
               /translation="MKTCNOYTLALGVRRNRPIMGVNGDQPPRRRROTROSEGIN
               MDNDVDEKTEKMEEDVDVDMNREKREYASGSSVDMNNAQRLKRHETTHHISM
               THMEIEVRLQKSKTDNMEAEIKERKRYRDLVIRLTAAYKTLAEQREKVDNDS
               GNVLAFIRMGFDLYMKDOIIRANKGEIHCFLSHKTONELTELLANRKMILLKI
               KDQKFSVTLDSPPDSRREKQMTFLRCYDVSTGSCRIEFPSTFLHDKGEGLEK
               ITLKLISQTRMESRDSLSKPIRYQAKRIDALFYLAESDNRHRSVDLSKESERTNC
               ICGFEFLFGWVLIWYDLAENVIVFPFKPRILVKKKKPFEGEDEKIDESTAKESFTIH
               YPMINDQALMSIEVFEQFOYEKIFGLFGVRLKVADEDELEKTSCKLKSASIKHD
               VHSVDVGEDLFMELKLDVLPKEIKTGPVLDPLKRMGCVPTNWTYRIILITPVS
               IASERTPEPKLILKNYLRSVPOERLNGALISIEOLYKKLDIOKLINERAGAKAR
               RIMFRHEKLYCIEFLITFLFNFELKONKHITLFCCKILGH"
               complement(36504..37124)
               /note="gb|AAF24530.1
               gene_id:MED5.9
               similar to unknown protein"
               /codon_start=1
               /evidence=not_experimental
               /protein_id="BAB01788.1"
               /db_xref="GI:9293885"
               /translation="MFRIVLAHACIFALISGDLLELTSLQADLITNNHSTRSSRS
               FRISDHSHTSRSGHHHTSPSLDEICPHLHOTITRSLHSTRSSAFETAPRL
               DLVDPGRGVSPSOHLTAVSMTSFPAFPHSTRSSRRKRRLOIFTRPLTRPGSS
               TVLNSKYFVVISIVSGYLAISSMYSSFQSTRSLPSFISADLVF"
               complement(38893..45012)
               /note="gene_id:MED5.10"
               /codon_start=1
               /pseudo
               /evidence=not_experimental
               /product="Athila retroelement ORF1 protein-like"
               complement(46183..46803)
               /note="gb|AAF24530.1
               gene_id:MED5.11

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 1, 2002, 20:11:08 ; Search time 159.19 Seconds
(without alignments)
4076.859 Million cell updates/sec

Title: US-09-586-106-62

Perfect score: 757
Sequence: 1 gaaagaaggtgtgaagctc.....atcaagaattcacaaggt 757

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_1101:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	381.6	50.4	163319	21	AAF22306 Arabidopsis thaliana
2	380	50.2	59590	21	AAF22281 BAC containing rep
3	375.2	49.6	79122	21	AAF22294 BAC containing rep
4	375.2	49.6	94895	21	AAF22302 BAC containing rep
5	375.2	49.6	94895	21	AAF22302 BAC containing rep
6	375.2	49.6	1082138	21	AAF22305 Arabidopsis thaliana
7	371.2	49.0	129021	21	AAF22296 BAC containing rep
8	370.4	48.9	103929	21	AAF22287 BAC containing rep
9	370.4	48.9	134499	21	AAF22286 BAC containing rep
10	366.4	48.4	109973	21	AAF22298 BAC containing rep
11	364.8	48.2	64415	21	AAF22279 BAC containing rep

12	357	47.2	12286	21	AAF23261 Plant generic retr
13	356	47.0	83390	21	AAF22283 BAC containing rep
14	356	47.0	90336	21	AAF22289 BAC containing rep
15	351.8	46.5	92584	21	AAF22288 BAC containing rep
16	349	46.1	4609	21	AAF23273 Soybean retroelement
17	348.6	46.1	9139	21	AAF23274 Soybean retroelement
18	344.8	45.5	611590	21	AAF22303 Arabidopsis thaliana
19	335.4	44.3	1082138	21	AAF22305 Arabidopsis thaliana
20	333.8	44.1	9829	21	AAF23271 Soybean retroelement
21	333	44.0	40349	21	AAF22278 Soybean retroelement
22	331.4	43.8	12571	21	AAF23272 Soybean retroelement
23	331.4	43.8	64415	21	AAF22279 BAC containing rep
24	316.6	41.8	96988	21	AAF22290 Arabidopsis retroe
25	309.4	40.9	600	21	AAF23284 BAC containing rep
26	302.4	39.9	80450	21	AAF22295 Plant retroelement
27	273.2	36.1	600	21	AAF23258 Soybean retroelement
28	259	34.2	597	21	AAF23280 Pea retroelement c
29	244.6	32.3	597	21	AAF23286 Soybean retroelement
30	222.6	29.4	109973	21	AAF22298 Soybean retroelement
31	214.6	28.3	10482	21	AAF23275 BAC containing rep
32	120.2	15.9	90336	21	AAF22284 Arabidopsis thaliana
33	120.2	15.9	121001	21	AAF22284 Arabidopsis thaliana
34	112.4	14.8	163319	21	AAF22306 BAC containing rep
35	108	14.3	82588	21	AAF22301 BAC containing rep
36	107.8	14.2	101786	21	AAF22293 BAC containing rep
37	105.2	13.9	8435	20	AAF23445 BAC containing rep
38	101.4	13.4	86584	21	AAF22292 Pepper Bst2 gene
39	97.6	12.9	31491	22	AAF2301 Pepper Bst2 resista
40	97.6	12.9	31491	22	AAF2301 BAC containing rep
41	93.2	12.3	101786	21	AAF22293 Oligonucleotide D1
42	91	12.0	936	22	AAF22293 Oligonucleotide D1
43	91	12.0	936	22	AAF22293 Oligonucleotide D1
44	91	12.0	936	22	AAF22293 Oligonucleotide D1
45	91	12.0	936	22	AAF22293 Oligonucleotide D2

ALIGNMENTS

RESULT 1

AAF22306 standard; DNA: 163319 BP.

ID	Score	Query Match	Length	DB ID	Description
1	381.6	50.4	163319	21	AAF22306 Arabidopsis thaliana
2	380	50.2	59590	21	AAF22281 BAC containing rep
3	375.2	49.6	79122	21	AAF22294 BAC containing rep
4	375.2	49.6	94895	21	AAF22302 BAC containing rep
5	375.2	49.6	94895	21	AAF22302 BAC containing rep
6	375.2	49.6	1082138	21	AAF22305 Arabidopsis thaliana
7	371.2	49.0	129021	21	AAF22296 BAC containing rep
8	370.4	48.9	103929	21	AAF22287 BAC containing rep
9	370.4	48.9	134499	21	AAF22286 BAC containing rep
10	366.4	48.4	109973	21	AAF22298 BAC containing rep
11	364.8	48.2	64415	21	AAF22279 BAC containing rep

17-MAR-2000; 2000WO-US07392.

18-MAR-1999; 99US-0125219.

01-APR-1999; 99US-0127409.

18-MAY-1999; 99US-0134770.

13-SEP-1999; 99US-0153584.

17-SEP-1999; 99US-0154603.

(UYCH-) UNIV CHICAGO.

Preuss D, Copenhaver G, Keith K.

WPI; 2000-587529/55.

Recombinant DNA construct comprising a plant centromere, useful for producing stably inherited microchromosomes which can serve as vectors for

CC selected proteins such as hormones, enzymes, interleukins, clotting

CC factors, cytokines, antibodies, and growth factors.

AA sequence 79122 BP; 23719 A; 15988 C; 16603 G; 22811 T; 1 other;

Query Match	49.6%	Score 375.2;	DB 21;	Length 79122;
Best Local Similarity	68.5%;	Pred. No. 4.2e-101;		
Matches 518; Conservative	0;	Mismatches 238;	Indels 0;	Gaps 0;

Oy	2	aaagaaggttttgaagccctcgggaaggaagtatattacatcaatcgtcgtccatctatgctgc	2
Db	67132	aaagaagaatttgaagaactgcgttgatgctcgatattcaattacccattcttcgtagaagtaactg	67191
Oy	62	ggtgagtcocggttgcataagcgtttcctcaagaagaagggagcattaccgttttcctcaatgttaa	121
Db	67192	ggttcctccagtgatcattgcattccctcaaaaaagcgagaatgactgctgttataaactgaaaa	67251
Oy	122	ggatgaattgattcccgacgaggaattatcaacacgtgcgcatagtgatgtgtatgcataattccagaa	181
Db	67252	ggatgagattgattcccaactagaactatacattacacgtggcagatagaattgtgtattataga	67311
Oy	182	actgaataagcgcattcaggaagaatcatcattacccttgcctttatcgacatatgcata	241
Db	67312	gcttaaatgtgtgcatctcaggaagaatcatcttcctcctacatcatcaatcaatcaatgcttga	67371
Oy	242	aaggtttgcacaactcacaacactctgcctttcttataagcgttatcttaattttccccaat	301
Db	67372	acgcttagcttaatcacccataactactacgctttcttcttgatgataagtggtttctttcaat	67431
Oy	302	accagttgcacaatctgcattcagggagaaacaccttcaacgtgccttccttcgtaacctgc	361
Db	67432	atcaattccccaatgatacgaagaagaacacaccttcaacatgctcttatgtaacctttgc	67491
Oy	362	ttaagacgtatgccttttgcgttatgttaatgtaacctgcaccttccaaagaatgataat	421
Db	67492	tctaaagaagaatgctcatttggttatgcaaatgccttcgaatatctcagaaggtgataac	67551
Oy	422	ggctataattctctaactcttgtgnaaaatattgtcgaggtttccatcagatgactttccgt	481
Db	67552	cctcatattttttattttaatcgaaggagatgtgtgaggtcttcatlgaatgatttttcagtt	67611
Oy	482	ttaaggtcttcttttggatgattgctctcgaacacgttgcagtttcgagtttcacagagatgtaa	541
Db	67612	ctatggccctcttcttcctcctcatgcttgtgaatccttgcgaagatattgaccaagttgcga	67671
Oy	542	agacaacaactctgtcttgaattgggagaaatgcacattatggttaatgaaagcaatcgt	601
Db	67672	aaagaaccaactcgttctcattatgggnaaaagtctcattcatggtgaaggaagcataagt	67731
Oy	602	cttggagacataaaattcttgaagaaggtatgaaatgcgataagcgaaggtttgataat	661
Db	67732	attgggtcacaaaggtattagagaaggttatagaaggttgcacatlggnaaaaaattaaatgtat	67791
Oy	662	cgaacaaatgcataaccacacagatatcaagttataagaaggtttccttggtaactgtgtg	721
Db	67792	gatcgaattgcacacacacaaaacctgtcnaaagacatcaagaagtttcccttggtaactgtg	67851
Oy	722	tttctatagaaggttatcaaaagtttccacaaggt	757
Db	67852	gtctccaggaagatttataagaagcttcccaagat	67887

	RESULT 4
AAE22302	
ID	AAE22302 standard; DNA; 94895 BP.
XX	
AC	
XX	AAE22302;
XX	
DT	20-MAR-2001 (first entry)
XX	
DE	BAC containing repeats from centromeres 1-4 #25
XX	
KW	Centromere; michrosome; vector; ds.

```

XX Arabidopsis thaliana.
OS
XX WO200055325-A2.
PN
XX 21-SEP-2000.
PD
XX 17-MAR-2000; 2000WO-US07392.
FE
XX 18-MAR-1999; 99US-0125219.
PR 01-APR-1999; 99US-0127409.
PR 18-MAY-1999; 99US-0134770.
PR 13-SEP-1999; 99US-0153584.
PR 17-SEP-1999; 99US-0154603.
XX (UYCH-) UNIV CHICAGO.
PA
XX Preuss D, Copenhaver G, Keith K;
PI
XX WPI; 2000-587529/55.
DR
XX Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited microsome which can serve as vectors for
PT the construction of transgenic plant and animal cells.
XX Claim 102; Page 804-819; 1449pp; English.
XX
XX The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited microsome which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX
XX Sequence 94895 BP; 28943 A; 18101 C; 18466 G; 29385 T; 0 other;
SQ

```

Query Match 49.6%; Score 375.2; DB 21; Length 94895;
 Best Local Similarity 68.5%; Pred. No. 4.6e-101;
 Matches 518; Conservative 0; Mismatches 238; Indels 0; Gaps 0;

```

QY 2 aaaaagggtgtgaaagctcctgtaagatattatctatcatgtgtgtccatagcattg 61
DB 45846 aaaaagaatttgaacatgctgagtgtgtcattcaacctcctcgtatgacttg 45905
QY 62 ggtgagtcctggtgcatagcgttccttaagaaggagcattacccgtgtccctaataa 121
DB 45906 ggttcttcacgtgcatgtcgtccctaaagggtggaatgactgtgtcacaataagaaa 45965
QY 122 ggaatgaatgacccgcagagatatactgctatagatgagtgtgcatgattcaga 181
DB 45966 ggaatgaatgacccgcactactagaacgtgcatgattgattgattgattgaa 46025
QY 182 actgaataaagccactagaagaagatcattaccccttcttcttcttcttcttcttct 241
DB 46026 gctaattgtcatctagagaagatcatttcttcttcttcttcttcttcttcttct 46085
QY 242 aaggtgtcccaactcaacacctccttcttcttcttcttcttcttcttcttcttct 301
DB 46086 aagcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 46145
QY 302 accgttccacatctgtatcagagaagaacacattacccgtcccttcttcttcttcttct 361
DB 46146 accaattcaccatctgtatcagagaagaacacattcaccatgtcttcttcttcttct 46205
QY 362 ttatagacgtatgcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 421
DB 46206 ttataaagaatgcatctgttcttcttcttcttcttcttcttcttcttcttcttct 46265
QY 422 ggcatacttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 481
DB 46266 ctcatacttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 46325

```

```

QY 482 ttaagggtcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 541
DB 46326 ctatggccctcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 46385
QY 542 agaccacatcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 601
DB 46386 ggaagcaactcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 46445
QY 602 cttaagacataaattcttcttcttcttcttcttcttcttcttcttcttcttcttct 661
DB 46446 gtgggttcaacaagatatacgaagaagggtatagaggttgcacaaaggaaatgaa 46505
QY 662 cgacaaatgcatcaccacagatatacaaggtatcaaggttcttcttcttcttcttct 721
DB 46506 gatgcaggtgcagccacagaaatgtgaaagacagaagaagttcttcttcttcttct 46565
QY 722 ttctataagaaggttatcaaaagattcaaaagt 757
DB 46566 gtctcacaagaattatataaagacttctcaaat 46601

```

RESULT 5

AAE22302/C

ID AAE22302 standard; DNA; 94895 BP.

XX AAE22302;

XX 20-MAR-2001 (first entry)

XX BAC containing repeats from centromeres 1-4 #25.

XX Centromere; microsome; vector; ds.

XX Arabidopsis thaliana.

XX WO200055325-A2.

XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-US07392.

XX 18-MAR-1999; 99US-0125219.

XX 01-APR-1999; 99US-0127409.

XX 18-MAY-1999; 99US-0134770.

XX 13-SEP-1999; 99US-0153584.

XX 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhaver G, Keith K;

XX WPI; 2000-587529/55.

XX Claim 102; Page 804-819; 1449pp; English.

The present invention relates to a recombinant DNA construct of a plant (Arabidopsis thaliana) centromere. The constructs are useful for producing stably inherited microsome which can serve as vectors for the construction of transgenic plant and animal cells expressing selected proteins such as hormones, enzymes, interleukins, clotting factors, cytokines, antibodies, and growth factors.

Sequence 94895 BP; 28943 A; 18101 C; 18466 G; 29385 T; 0 other;

Query Match 49.6%; Score 375.2; DB 21; Length 94895;
 Best Local Similarity 68.5%; Pred. No. 4.6e-101;
 Matches 518; Conservative 0; Mismatches 238; Indels 0; Gaps 0;

XX WPI: 2000-587529/55.
 XX Recombinant DNA construct comprising a plant centromere, useful for
 PT producing stably inherited microsomemes which can serve as vectors for
 PT the construction of transgenic plant and animal cells
 XX
 PS Claim 102: Page 484-508; 1449pp; English.
 CC The present invention relates to a recombinant DNA construct of a plant
 CC (Arabidopsis thaliana) centromere. The constructs are useful for
 CC producing stably inherited microsomemes which can serve as vectors for
 CC the construction of transgenic plant and animal cells expressing
 CC selected proteins such as hormones, enzymes, interleukins, clotting
 CC factors, cytokines, antibodies, and growth factors.
 XX
 SQ Sequence 103929 BP; 32987 A; 19310 C; 18547 G; 33085 T; 0 other;

Query Match 48.9%; Score 370.4; DB 21; Length 103929;
 Best Local Similarity 68.1%; Pred. No. 1.3e-99;
 Matches 515; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

OY 2 aaaaagagttgtgaagctccctgagagagttatctatcatgtgtcctatagagatg 61
 DB 18128 AAGAGGATGATCTTAACTATGATGCTGGGTTATCTACCTATCTGATGACACTTG 18069
 OY 62 ggtgagtcggtgcatagcgttcctaaagaggagcattaccgtgtccctaataa 121
 DB 18068 GGTATCTCCAGTTCACTGGCTTCTTAAAGAGGATGACAGTGTAAATTTCTAA 18009
 OY 122 ggaatgatccgcagagagattatcaactgctatagatggtgatgttcaagaa 181
 DB 18008 AGATGAACGATGACACCTATGATGATGATGATGATGATGATGATGATGATGAT 17949
 OY 182 acggaataagagcagagagagagagagagagagagagagagagagagagag 241
 DB 17948 GTTAAATGTTGCTCTAGAAAGAACATTTCCATTTGCCCTTATGATCATATGCTA 17889
 OY 242 aaggtgtccaaactacacactctgcttcttagaaggttattctagttctccaa 301
 DB 17888 AATATTGACAAACCATCTTACTATGCTTTTGTATGATGATGATGATGATGATGAT 17829
 OY 302 accagttgcaaatctgcatagagagagagagagagagagagagagagagag 361
 DB 17828 CCTATTCACCCAAATGACCAAGAAACCACTTTCATTTGCTTATGAGACTTTCG 17769
 OY 362 ttatagacgtatgctcttctgctatgtaagtcacgtccacacttcaagaatgta 421
 DB 17768 TTACAGACGATATGCTTTCGCTCTCTGTATGACACCAAGCTTTCAGCGGTATGAC 17709
 OY 422 ggcataatctctaaacttctggaataatctcgaggttctatgagatcttccgt 481
 DB 17708 TTCTATTTTCGATTTGATGATGAGGATGATGATGATGATGATGATGATGATGAT 17649
 OY 482 ttaaggtctctcttctgagatgattgctccagcaactctgacgtatcttaagaatg 541
 DB 17648 GTATGAGCTCTTCT 17589
 OY 542 agacacaaatctgcttctgagagagagagagagagagagagagagagagagag 601
 DB 17588 AGAAGACAACTGCTGTAACCTGGAGAAATGCAATTCATGCTGATGAGAAAGCATG 17529
 OY 602 cttagagataaaatctctgagaagagatgagatgagatgagatgagatgagat 661
 DB 17528 TTTGGGCGCGGAAATTTCTGAGGAGGATGATGATGATGATGATGATGATGATGAT 17469
 OY 662 cgacaaatgcatccacacagatatacaagagatagagagagagagagagagagag 721
 DB 17468 GATGACGTTACAGCACCACCAAAATGCTCAAGAGATCAGAGGTTTCTTGACATG 17409
 OY 722 ttctatagaggttattcaagaattcaagaagt 757

DB 17408 ATTTTACAGATATTCATCAGAGATTTCTCCAGT 17373

RESULT 9
 AAF22286/c
 ID AAF22286 standard; DNA; 134499 BP.
 XX
 AC AAF22286;
 XX
 DT 20-MAR-2001 (first entry)
 XX
 DE BAC containing repeats from centromeres 1-4 #9.
 XX
 KW Centromere; microsome; vector; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200055325-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 17-MAR-2000; 2000WO-US07392.
 XX
 PR 18-MAR-1999; 99US-0125219.
 PR 01-APR-1999; 99US-0127409.
 PR 18-MAY-1999; 99US-0134770.
 PR 13-SEP-1999; 99US-0153584.
 PR 17-SEP-1999; 99US-0154603.
 XX
 PA (UYCH-) UNIV CHICAGO.
 XX
 PI Preuss D, Copenhaver G, Keith K;
 XX
 DR WPI: 2000-587529/55.
 XX
 PT Recombinant DNA construct comprising a plant centromere, useful for
 PT producing stably inherited microsomemes which can serve as vectors for
 PT the construction of transgenic plant and animal cells
 XX
 PS Claim 102: Page 453-484; 1449pp; English.
 CC The present invention relates to a recombinant DNA construct of a plant
 CC (Arabidopsis thaliana) centromere. The constructs are useful for
 CC producing stably inherited microsomemes which can serve as vectors for
 CC the construction of transgenic plant and animal cells expressing
 CC selected proteins such as hormones, enzymes, interleukins, clotting
 CC factors, cytokines, antibodies, and growth factors.
 XX
 SQ Sequence 134499 BP; 41565 A; 25130 C; 25225 G; 42577 T; 2 other;

Query Match 48.9%; Score 370.4; DB 21; Length 134499;
 Best Local Similarity 68.1%; Pred. No. 1.4e-99;
 Matches 515; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

OY 2 aaaaagagttgtgaagctccctgagagagttatctatcatgtgtcctatagagatg 61
 DB 130511 AAGAGGATGATCTTAACTATGATGCTGGGTTATCTACCTATCTGATGACACTTG 130452
 OY 62 ggtgagtcggtgcatagcgttcctaaagaggagcattaccgtgtccctaataa 121
 DB 130451 GGTATCTCCAGTTCACTGCTTCTTAAAGAGGATGACAGTGTAAATTTCTAA 130392
 OY 122 ggaatgatccgcagagagattatcaactgctatagatggtgatgttcaagaa 181
 DB 130391 AGATGAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 130332
 OY 182 actgaataagagcag 241
 DB 130331 GTTAAATGTTGCTCTAGAAAGAACATTTCCATTTGCCCTTATGATCATATGCTA 130272
 OY 242 aaggtgtccaaactacacactctgcttcttagaaggttattctagttctccaa 301

RESULT 10
 AAF22298/C
 ID AAF22298 standard; DNA; 109973 BP.
 XX
 AC AAF22298;
 XX
 DT 20-MAR-2001 (first entry)
 XX
 DE BAC containing repeats from centromeres 1-4 #21.
 XX
 KW Centromere; microsome; vector; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN W0200055325-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 17-MAR-2000; 2000WO-US07392.
 XX
 PR 18-MAR-1999; 99US-0125219.
 PR 01-APR-1999; 99US-0127409.
 PR 18-MAY-1999; 99US-0134770.
 PR 13-SEP-1999; 99US-0153584.
 PR 17-SEP-1999; 99US-0154603.
 XX
 PA (UYCH-) UNIV CHICAGO.
 XX
 PI Preuss D, Copenhaver G, Keith K;
 XX
 DR WPI; 2000-587529/55.
 XX
 PT Recombinant DNA construct comprising a plant centromere, useful for
 PT producing stably inherited microsomes which can serve as vectors for
 PT the construction of transgenic plant and animal cells
 XX
 CS Claim 102; Page 738-763; 1443pp. English.

Query Match	48.4%;	Score 366.4;	DB 21;	Length 109973;
Best Local Similarity	68.0%;	Pred. No. 2.1e-98;		
Matches 511; Conservative	0;	Mismatches 241;	Indels 0;	Gaps 0

QY	3	aaagaggttgtgaagctcctggaatgaaggtatatactatcatatgcttgctcatagcattg	62
Db	93313	AAGGAATATTATGAACCTGTCCGGAAGAGAAATTATCTATCCAAATTTCAGATACAGCTGG	9325
QY	63	gtgagtcgggtgcatagcgttccttaagaaggagggcattacgctgtccctaattgaag	122
Db	93253	GTTTAGCGAGTTTCATCTAATTCTCTTAAGAAAGAGAGATTACAGTACTACTCAAGATGAGAA	9319
QY	123	gatgaattgatacccgcaagaagatatacctatggtgctatagatgtgtatgcatttcaagaa	182
Db	93193	GACGAGCTTAATCTTACTACCGCAACATCACAGGTCATCAATGTGCATTCGATTATAGAAAG	9313
QY	183	ctggaataaagccaactagaagaagaatcaattaccccttgcccttatacgacatatgtagaa	242
Db	93133	CTGAATGTGTGCATCACATGAAACACATTTCCCTTACCCTTTATTTAGTATCAGATGTGGAG	9307
QY	243	aggtgtccaaactacacaacactcgcgtcttcagaaggtatctcagttctcccaataa	302
Db	93073	AGGTTAGCAAAATCATACGACTATTGCTTCTTGATGGATACTCAGATCTTTTAGAAC	93014
QY	303	ccaagttgacaactatgatcatgagaagaanaaacacttaccctgccttcgtaacctgct	362
Db	93013	CCGATTTCATCCAGATGACCAGAGAAACGACTTTCACCTGCCCTATNGTACATTTCGT	92954
QY	363	tatagaactatgaccttgggttgtaatagtacatgcactgcacccttcaagaatgtatg	422
Db	92953	TATCGAGAAATGCCCTTGCGTCTTTGTAATCTCCTCAACATTTACAGAGAGTATGATG	92894
QY	423	gctatatctctaactcttctggaataatgtcgaaggttccataggaagactttcgtt	482
Db	92893	TCTATCTTCACAGACATGATTTGAGGATATCATGGAAGTTTTCATGAGATTTTTTAAT	92834
QY	483	taagggtctcttcttgatgatgtgcctcaagcaacctltgacgagcttcaagagatgtaaa	542
Db	92833	TATGTGATCATCTGTTTGAGGATGCTTGAGAGATCTCTACAAAGTGTGGCAAGATGTGAG	92774
QY	543	gaccccaactctgtccttgaataggggagaagaagtgccacttattggtttaaaggaacatg	602
Db	92773	GAGAAACATCTAGTTTGGAAATGGAAATAAAGTCACTTATAGTGCGAGATGACATTAAT	92714
QY	603	ttaagacataaaattctctgaagaaggtcatlgaagtcgataaaggtatgtagtgcacatc	662
Db	92713	CTCGGACACAGAAATTTGTGACATGTTATAGAACTTATAGCTTAATATGCAAGTATAG	92654
QY	663	gacaaaaatgcataaccccacagatatcaaaagttaaagttacagtttctctgtgtcatggtg	722
Db	92653	ACAAAGTCTTCAAGGCGCTGATATGTGTAAAGCATGTGAGGATTTCCCTGGCAATGCGTGT	92594
QY	723	ttctatagaaggttatacaagaatttcaaaa	754
Db	92593	TTCTACAGGAGATTTCATCAAAAGCTTTAGCAA	92562

RESULT	11
AAAF22279	
ID	AAAF22279 standard; DNA; 64415 BP
XX	
AC	AAAF22279;
XX	

[illegible]

QY	422	ggtcatattctctaactctgttgaanaatatgtctcgaggttttcatatgagatcttcggt	481
Db	60359	ctctaatatttcagacttaatcgaaggagatggttgaggattttcatatgatatcttcggt	60418
QY	482	tttcgggtcttccttttgatgatcttgccctcagcaaccttgatcgagctcttaagaagatlaa	541
Db	60419	ctatggccccctcttcctgcgtcatctgtttgtaaccttcggcagggttatgtactagtgcgga	60478
QY	542	agcacacaactctgtcttgaatgagggaagatgcccacttaatggtlaatgaagcatcgt	601
Db	60479	aggagcagcaactctgtctcctaatttggaaaagtgtcattcatatggtgaaggaaagcataagt	60538
QY	602	cttaggacataaattcttcgaagaaggtatgtgaagtcgataaggctcgaagttgatcaat	661
Db	60539	gcttgatcacacaagatatacagaagaagggtatcagaagtttgacaagaagaagttaagtgt	60598
QY	662	cgacaacaatcgccatacccaacagatatacgaagtaagaagtcttcctgtgcatagtggt	721
Db	60599	gatgcagatctgcagcaccaaaaacggtgaaggacatacgaagaatttcctgtgcatatgctgg	60658
QY	722	tttctataagaagtttatcaagaatttcacaagaagt	757
Db	60659	gtctcacagaagaatttatataaagactctcccaaat	60694
Db	60659	gtctcacagaagaatttatataaagactctcccaaat	60694
RESULT	12		
AAZ35261			
ID	AAZ35261	standard; DNA; 12286 BP.	
XX			
AC	AAZ35261;		
XX			
DT	27-MAR-2000	(first entry)	
XX			
DE	Plant generic retroelement.		
XX			
KW	Retroelement; retrovirus; transgenic plant; gene transfer;		
KW	soybean; pea; Calypso; athila; cyclops; ss.		
XX			
OS	Glycine max.		
OS	Arabidopsis thaliana.		
OS	Plisum sativum.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1482..6887	
FT		/*tag= a	
XX			
PN	W09960842-A2.		
XX			
PD	02-DEC-1999.		
XX			
PF	28-MAY-1999;	99WO-US11858.	
XX			
PR	29-MAY-1998;	98US-0087125.	
PR	28-MAY-1999;	99US-0087125.	
XX			
XX			
PA	(WRIG/) WRIGHT D A.		
PA	(VOYT/) VOYTAS D F.		
XX			
PI	Wright DA, Voytas DF;		
XX			
XX			
DR	WPI: 2000-105586/09.		
XX	P-PSDB: AAY32434.		
XX			
PT	New nucleic acid molecules for imparting agronomically significant		
PT	characters to plants, especially soybean		
XX			
PS	Claim 1(h): Page 84-88; 118pp; English.		
XX			
CC	The present sequence comprises a generic plant retroelement		
CC	obtained from retrovirus-like elements (retroelements) calypso of		
CC	soybean, cyclops of pea and athila of Arabidopsis thaliana. The		
CC	invention provides molecular tools in the form of retroelements and		

CC retroelement-containing vectors, cells and plants. Methods are
 CC provided for introducing the retroelements into cells, especially
 CC when the retroelement carries at least 1 agronomically-significant
 CC characteristic (ACS). In a preferred method, a helper cell line
 CC which expresses gag, pol and env sequences is used to enable
 CC transfer of a secondary construct which carries an ASC and has
 CC retroelement sequences that allow for replication and integration.
 CC Claimed isolated nucleic acid molecules comprise a nucleic acid
 CC sequence selected from a retroelement primer binding site,
 CC envelope, gag, integrase, reverse transcriptase, protease or
 CC RNase-H sequence (see AA35254-61). Also provided are plant
 CC retroviral particles that are used to transfer the nucleic acids
 CC into plant cells.

XX
 XX Sequence 12286 BP; 3748 A; 2540 C; 2767 G; 3231 T; 0 other;

Query Match 47.2%; Score 357; DB 21; Length 12286;
 Best Local Similarity 67.0%; Pred. No. 5e-96;
 Matches 507; Conservative 0; Mismatches 250; Indels 0; Gaps 0;

QY 1 gaaaagaggttgtaagctcctgtaaggtatattctatctatgttgcctatagcatt 60
 DB 4258 gaagagaggtgtaagctcctgtaaggtatattctatctatgttgcctatagcatt 60
 QY 61 ggggtggtcgggtgtaagctcctgtaaggtatattctatctatgttgcctatagcatt 120
 DB 4318 ggggtggtcgggtgtaagctcctgtaaggtatattctatctatgttgcctatagcatt 120
 QY 121 agagttaattgatactcctgtaaggtatattctatctatgttgcctatagcatt 180
 DB 4378 ggaatgacttgatactcctgtaaggtatattctatctatgttgcctatagcatt 180
 QY 181 aactgataaagcactgtaaggtatattctatctatgttgcctatagcatt 240
 DB 4438 agctgaatgataaagcactgtaaggtatattctatctatgttgcctatagcatt 240
 QY 241 aaggttgccaaactcctgtaaggtatattctatctatgttgcctatagcatt 300
 DB 4498 aaggttgccaaactcctgtaaggtatattctatctatgttgcctatagcatt 300
 QY 301 taecaggtgcaaatcctgtaaggtatattctatctatgttgcctatagcatt 360
 DB 4558 tcgaggtgcaaatcctgtaaggtatattctatctatgttgcctatagcatt 360
 QY 361 ctatagacgtatgcttctgtaaggtatattctatctatgttgcctatagcatt 420
 DB 4618 ctatagacgtatgcttctgtaaggtatattctatctatgttgcctatagcatt 420
 QY 421 tggcctatctcctatgtaaggtatattctatctatgttgcctatagcatt 480
 DB 4678 tggcctatctcctatgtaaggtatattctatctatgttgcctatagcatt 480
 QY 481 tttaaggtctcctatgtaaggtatattctatctatgttgcctatagcatt 540
 DB 4738 tttaaggtctcctatgtaaggtatattctatctatgttgcctatagcatt 540
 QY 541 aagacacaaatctgtaaggtatattctatctatgttgcctatagcatt 600
 DB 4798 aagacacaaatctgtaaggtatattctatctatgttgcctatagcatt 600
 QY 601 tcttagagcacaataattctgtaaggtatattctatctatgttgcctatagcatt 660
 DB 4858 tcttagagcacaataattctgtaaggtatattctatctatgttgcctatagcatt 660
 QY 661 tcgacaaaatgcaatcctgtaaggtatattctatctatgttgcctatagcatt 720
 DB 4918 tcgacaaaatgcaatcctgtaaggtatattctatctatgttgcctatagcatt 720
 QY 721 gtctctagaaggttatacaaatctcacaaggt 757
 DB 4978 gtctctagaaggttatacaaatctcacaaggt 757

RESULT 13
 AAF22283
 ID AAF22283 standard; DNA; 83390 BP.
 AC
 XX
 AC AAF22283;
 XX
 DT 20-MAR-2001 (first entry)
 XX
 DE BAC containing repeats from centromeres 1-4 #6.
 XX
 KM Centromere; microsome; vector; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN W020005325-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 17-MAR-2000; 2000MO-US07392.
 XX
 PR 18-MAR-1999; 99US-0125219.
 PR 01-APR-1999; 99US-0127409.
 PR 18-MAY-1999; 99US-0134770.
 PR 13-SEP-1999; 99US-0153584.
 PR 17-SEP-1999; 99US-0154603.
 XX
 PA (UYCH-) UNIV CHICAGO.
 XX
 PI Preuss D, Copenhaver G, Keith K;
 XX
 DR WPI; 2000-587529/55.
 XX
 PT Recombinant DNA construct comprising a plant centromere, useful for
 PT producing stably inherited microsome which can serve as vectors for
 PT the construction of transgenic plant and animal cells

XX
 PS Claim 102; Page 386-404; 1449pp; English.
 XX
 CC The present invention relates to a recombinant DNA construct of a plant
 CC (Arabidopsis thaliana) centromere. The constructs are useful for
 CC producing stably inherited microsome which can serve as vectors for
 CC the construction of transgenic plant and animal cells expressing
 CC selected proteins such as hormones, enzymes, interleukins, clotting
 CC factors, cytokines, antibodies, and growth factors.
 XX

SO Sequence 83390 BP; 24664 A; 17305 C; 17224 G; 24140 T; 57 other;

Query Match 47.0%; Score 356; DB 21; Length 83390;
 Best Local Similarity 66.9%; Pred. No. 2.3e-95;
 Matches 506; Conservative 0; Mismatches 250; Indels 0; Gaps 0;

QY 2 aaaaagaggttgtaagctcctgtaaggtatattctatctatgttgcctatagcatt 61
 DB 45441 aaaaagaggttgtaagctcctgtaaggtatattctatctatgttgcctatagcatt 61
 QY 62 ggtgagtcgggtgtaagctcctgtaaggtatattctatctatgttgcctatagcatt 121
 DB 45501 ggtgagtcgggtgtaagctcctgtaaggtatattctatctatgttgcctatagcatt 121
 QY 122 ggaatgatactcctgtaaggtatattctatctatgttgcctatagcatt 181
 DB 45561 ggaatgatactcctgtaaggtatattctatctatgttgcctatagcatt 181
 QY 182 actgataaagcactgtaaggtatattctatctatgttgcctatagcatt 241
 DB 45621 actgataaagcactgtaaggtatattctatctatgttgcctatagcatt 241
 QY 242 aaggttgccaaactcctgtaaggtatattctatctatgttgcctatagcatt 301
 DB 45681 aaggttgccaaactcctgtaaggtatattctatctatgttgcctatagcatt 301

BAC containing repeats from centromeres 1-4 #11.

KW Centromere; microsome; vector; ds.

OS *Arabidopsis thaliana*.

PN W0200055325-A2.

PD 21-SEP-2000.

PF 17-MAR-2000; 2000WO-US07392.

PR 18-MAR-1999; 99US-0125219.

PR 18-MAY-1999; 99US-0134770.

PR 17-SEP-1999; 99US-0154603.

PA (UYCH-) UNIV CHICAGO.

25	Preuss D,	Copenhagen G,	Ke
PI			

WPI; 2000-587529/55.

Recombinant DNA construct comprising

PT producing stably inherited microsomes which can serve as vectors for the construction of transgenic plant and animal cells -

Claim 102; Page 508-529; 1449pp; English

CC The present invention relates to a recombinant DNA construct of a plant
CC (*Arabidopsis thaliana*) centromere. The constructs are useful for
CC producing stably inherited mitochondria which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.

S0 Sequence 92584 BP; 27840 A; 18113 C; 19835 G; 26796 T; 0 other;

2000

Model	Best Local Similarity	Pred. No.
Model 1	68.38	4.3e-94
Model 2	68.38	4.3e-94
Model 3	68.38	4.3e-94
Model 4	68.38	4.3e-94
Model 5	68.38	4.3e-94
Model 6	68.38	4.3e-94
Model 7	68.38	4.3e-94
Model 8	68.38	4.3e-94
Model 9	68.38	4.3e-94
Model 10	68.38	4.3e-94
Model 11	68.38	4.3e-94
Model 12	68.38	4.3e-94
Model 13	68.38	4.3e-94
Model 14	68.38	4.3e-94
Model 15	68.38	4.3e-94
Model 16	68.38	4.3e-94
Model 17	68.38	4.3e-94
Model 18	68.38	4.3e-94
Model 19	68.38	4.3e-94
Model 20	68.38	4.3e-94
Model 21	68.38	4.3e-94
Model 22	68.38	4.3e-94
Model 23	68.38	4.3e-94
Model 24	68.38	4.3e-94
Model 25	68.38	4.3e-94
Model 26	68.38	4.3e-94
Model 27	68.38	4.3e-94
Model 28	68.38	4.3e-94
Model 29	68.38	4.3e-94
Model 30	68.38	4.3e-94
Model 31	68.38	4.3e-94
Model 32	68.38	4.3e-94
Model 33	68.38	4.3e-94
Model 34	68.38	4.3e-94
Model 35	68.38	4.3e-94
Model 36	68.38	4.3e-94
Model 37	68.38	4.3e-94
Model 38	68.38	4.3e-94
Model 39	68.38	4.3e-94
Model 40	68.38	4.3e-94
Model 41	68.38	4.3e-94
Model 42	68.38	4.3e-94
Model 43	68.38	4.3e-94
Model 44	68.38	4.3e-94
Model 45	68.38	4.3e-94
Model 46	68.38	4.3e-94
Model 47	68.38	4.3e-94
Model 48	68.38	4.3e-94
Model 49	68.38	4.3e-94
Model 50	68.38	4.3e-94
Model 51	68.38	4.3e-94
Model 52	68.38	4.3e-94
Model 53	68.38	4.3e-94
Model 54	68.38	4.3e-94
Model 55	68.38	4.3e-94
Model 56	68.38	4.3e-94
Model 57	68.38	4.3e-94
Model 58	68.38	4.3e-94
Model 59	68.38	4.3e-94
Model 60	68.38	4.3e-94
Model 61	68.38	4.3e-94
Model 62	68.38	4.3e-94
Model 63	68.38	4.3e-94
Model 64	68.38	4.3e-94
Model 65	68.38	4.3e-94
Model 66	68.38	4.3e-94
Model 67	68.38	4.3e-94
Model 68	68.38	4.3e-94
Model 69	68.38	4.3e-94
Model 70	68.38	4.3e-94
Model 71	68.38	4.3e-94
Model 72	68.38	4.3e-94
Model 73	68.38	4.3e-94
Model 74	68.38	4.3e-94
Model 75	68.38	4.3e-94
Model 76	68.38	4.3e-94
Model 77	68.38	4.3e-94
Model 78	68.38	4.3e-94
Model 79	68.38	4.3e-94
Model 80	68.38	4.3e-94
Model 81	68.38	4.3e-94
Model 82	68.38	4.3e-94
Model 83	68.38	4.3e-94
Model 84	68.38	4.3e-94
Model 85	68.38	4.3e-94
Model 86	68.38	4.3e-94
Model 87	68.38	4.3e-94
Model 88	68.38	4.3e-94
Model 89	68.38	4.3e-94
Model 90	68.38	4.3e-94
Model 91	68.38	4.3e-94
Model 92	68.38	4.3e-94
Model 93	68.38	4.3e-94
Model 94	68.38	4.3e-94
Model 95	68.38	4.3e-94
Model 96	68.38	4.3e-94
Model 97	68.38	4.3e-94
Model 98	68.38	4.3e-94
Model 99	68.38	4.3e-94

QY	3	aagaagcttgcgaagcccttcgataaaggtatctacttcacgttgcctcatgacatg	62
Db	79317	aagaagatcatgaagatgtctggaaagtctgggacatcctatccatttaagacagatgatg	79376
QY	63	gtgaagctccgylgcalaagcttccctaagaagaggagacatccgttcgctccatgataag	122
Db	79377	gttagctcagcgacatgigtctccctaagaagaagagtglttagagtggtttaagaatgaca-	79435
QY	123	gatgaattgatgcccgacagaagattatcaactggtcctataagatggtgacatgcattcagaa	182
Db	79436	aatgaagctgattcctcatagcgactgltcccgacagacagatgtagtgaactgcaggaag	79495
QY	183	ctgaataaagccactataagaaagacatcaacccttgcctttatagacatatgttagaa	242
Db	79496	ctgaagcctgcacaagaagaaatccatttcccttaacatcatatgatacactgttgaa	79555
QY	243	aggttgtccaaactacacacactctgcttctctagaagcttactctagttctcccaata	302
Db	79556	aggttactatcaacaaagtactactgctgtctgtatgtagtactcaaggtctcttcagata	79615
QY	303	ccagttgcacaacttgatccagagaaaaacacacttcaacctgcctttcgttaacctgtc	362
Db	79616	ccaatccatccagaatgacgaagaaagacacatgttcaacgtttccctatgatactttgct	79675
QY	363	tatagacgtatgacctttgctgttgaatgaatgacactgcacacttccaagaagttagatg	422
Db	79676	tatagcagaatgacattggtgtctatgtgcaatgctccctgatatatttgagaagatgtatgacg	79735
QY	423	gctatatctctctaaacttttgbaaataattgtcgaggttccatgatagatacttccgt	482

Db 79736 tcaatctcacagacatgatctgagaattccatagaaagctcttatgatctgtcttctcgt 797359
QY 483 tacggtctcttttttgatgatctgctcaagcaaccttgatcgagctcttacagagtgtaa 542
Db 79796 tatggtctctcatcttgaacgctgcttgagaaaccttcgcaaaagctctagctagatgtga 798555
QY 543 gacaccaactctgtcttgatgtggagaaagtcaccttatgtgttaatgaagcctgc 602
Db 79856 gaggagaatttgctctctaatatgggagaaagttcatcttcacgctcaaggaaggtatagt 799151
QY 603 tttaggacataaaattctctgaagaagctatctgaagtcagataagcgtaagttatcaatc 662
Db 79916 ctgtgtccacaagttctcagtgctgtgtatgttggtataaagaagccaagtaagttatc - 79974
QY 663 gacaaatgcccatacccacagatatcaaaagtataagaagttctcctgtcatgtgtgt 722
Db 79975 gacaagttcacgagcactcagctagatgtgaatctgtgagagattctcttgacatcttga 80034
QY 723 ttctatagaagttatacaaaagattccaaggt 757
Db 80035 ttctacgagaggttcgtcaagatcttcagtaagat 80066

```
Search completed: April 1, 2002, 22:50:15
Job time: 9547 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 1, 2002, 20:03:58 ; Search time 1412.44 Seconds
(without alignments)
5759.221 Million cell updates/sec

Title: US-09-586-106-62

Perfect score: 757
Sequence: 1 gaaagaaggtgtgaagctc.....atcaagaattcacaaagt 757

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_estin:*
3: em_estin:*
4: em_estin:*
5: em_estin:*
6: em_estin:*
7: em_estin:*
8: em_estin:*
9: em_estin:*
10: em_estin:*
11: em_estin:*
12: em_estin:*
13: em_estin:*
14: em_estin:*
15: em_estin:*
16: em_estin:*
17: em_estin:*
18: em_estin:*
19: em_estin:*
20: em_estin:*
21: em_estin:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	303.8	40.1	667	13	AO956471 LERAK19TR
2	288.4	39.4	680	13	B67757 T24P14NF TA
3	288.8	38.2	624	13	AO956470 LERAK19TR
4	266.2	35.2	660	13	AO967851 LERAK19TR
5	262.8	34.7	622	13	B26562 T10A4TR TAM
6	259	34.2	720	13	CNS005WS Arabidops
7	252.8	33.4	618	13	AO966890 LERAK19TR
8	243	32.1	612	13	AO368125 LERAK19TR
9	239.8	31.7	746	13	AO960305 LERAK19TR
10	235.4	30.1	617	13	AO368305 LERAK19TR
11	227.8	29.4	510	13	BF657025 OV2_18_CO
12	222.2	29.4	510	13	AO966889 LERAK19TR

13	221.2	29.2	726	13	AO958534 LERAK19TR
14	214.8	28.4	510	13	AO967852 LERAK19TR
15	213	28.1	785	13	BH023426 GH_MBO000
16	212	28.0	699	13	BH012654 TDGAM25TH
17	206.8	27.3	694	13	AO956739 LERAK19TR
18	206.2	27.2	514	13	AO966888 LERAK19TR
19	205	27.1	629	13	AO688565 LERAK19TR
20	202.4	26.7	453	13	BH009070 LERAK19TR
21	194.6	25.7	412	13	CNS00YKB
22	194.6	25.7	737	13	BH013137 LERAK19TR
23	192.8	25.5	645	13	AO367606 LERAK19TR
24	192.6	25.4	617	13	AO327960 LERAK19TR
25	182.2	24.1	649	13	BH011999 LERAK19TR
26	177.8	23.5	524	13	BH143723 LERAK19TR
27	176	23.2	561	13	B27354 LERAK19TR
28	175.6	23.2	640	13	BH007538 LERAK19TR
29	171.6	22.7	608	13	AO355556 LERAK19TR
30	168.8	22.3	412	13	BH009071 LERAK19TR
31	164.4	21.7	885	13	AO861394 LERAK19TR
32	153.6	20.3	736	13	BH022333 LERAK19TR
33	153.4	20.3	533	13	B27662 LERAK19TR
34	149.6	19.8	401	13	B29704 LERAK19TR
35	147.6	19.5	499	13	CNS00XWC
36	141.6	18.7	534	13	CNS00U91
37	140.6	18.6	825	13	B19887 LERAK19TR
38	132.8	17.5	697	13	AO960306 LERAK19TR
39	127.8	16.9	352	13	CNS00YCA
40	127	16.8	489	13	CNS00S9W
41	126.8	16.8	410	13	B27967 LERAK19TR
42	126	16.6	588	13	AO273556 LERAK19TR
43	125	16.5	693	13	AO689636 LERAK19TR
44	124	16.4	1062	13	CNS005AL
45	121.8	16.1	614	13	A2044562 Gm_UMB001

ALIGNMENTS

RESULT 1
LOCUS AO956471 667 bp DNA
DEFINITION LERAK19TR LERA Arabidopsis thaliana genomic clone LERAK19, DNA
ACCESSION AO956471 GI:6784172
VERSION AO956471.1
KEYWORDS
SOURCE GSS
ORGANISM thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidops.
1 (bases 1 to 667)
Buell, C.R., Lin, X., Pal, G., Barnstead, M., Bowman, C., Ullrich, T.,
Felbljum, T., Liang, F., Creasy, T. and Fraser, C.M.
Genomic survey sequencing of Landsberg erecta ecotype of
Arabidopsis thaliana and identification of sequence-based
polymorphisms
Unpublished (2000)
Contact: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: atc@igir.org
For additional information, see <http://www.tigr.org/tdb/at/at.html>
Seq primer: TR
Class: Shotgun.

FEATURES

source

Location/Qualifiers
1..667
/organism="Arabidopsis thaliana"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
/clone="LERAK19"

/clone_11b="LERA"
 /note="Organ: Leaf; Vector: PHOS1; Total genomic DNA was
 sheared to 0.9-1 kbp before ligation."
 BASE COUNT 203 a 111 c 160 g 192 t 1 others
 ORIGIN

Query Match 40.1%; Score 303.8; DB 13; Length 667;
 Best Local Similarity 69.3%; Pred. No. 7.6e-73;
 Matches 413; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

QY 3 aaagaggtgtgaagcttcctgagatgaagttatctatcatctgttcagatgagtg 62
 DB 71 AAAAGATTTATGAAACTGTTGGAGAGGAGATATCTATCCAAATGTTAGACAGTTAG 130
 QY 63 gtgagccggtgcatagcgttcctaaagaaggaagcattaccgtgtcccaatgataag 122
 DB 131 GTTAGTCCAGTTCTATGTGGTTCCTTAAGAGAGGAGTGTACAGTACAAAGATGAGAAA 190
 QY 123 gatgaattgattcccgagagatattacactgagatagatgagtgatcagattcagaaa 182
 DB 191 GAGGAGCTGATTCCTACTCGAGACATATAGACATGCGATGTCATCGATTACAGAAAG 250
 QY 183 ctgaatgaagcactaggaagatcattacccttgcctttatcgaccatattgtagaa 242
 DB 251 CTGAATGCTCTACAGAGAAAGACATTTCCCTTACCATTTATCGATGATGTTGAG 310
 QY 243 aggtgtccaaactcaacacctctgtcttagagaggttatctatcttcccaata 302
 DB 311 AGGTTAGCAATATATAGTACTATGCTTCTCTTATGATGATCTCAAAATGTTTCAATC 370
 QY 303 cgaattgacacatctgacagagaagaaacaccttcaacctgccttcggttaccctgtc 362
 DB 371 CCCAATCATCCAAATGACCGAGAGAAACGACTTTCACCTGCCCCATGATGATGATGCT 430
 QY 363 tatagagctatgctctgtgcttatgtatgaacgtgcacaccttcaagaatgataag 422
 DB 431 TATCAGAGATGCTCTGCTGCTTTGTAATGCTCTGCGCATTTTCAGAGAGATGATG 490
 QY 423 gctatactctcaactctgtgagaataatgtcgaagtttcaatgagttacccttccgt 482
 DB 491 TCTATCTTCACAGATATGATCGAGATATCATGAGAGTTTCAATGATGCTTTTCACT 550
 QY 483 taagagctctcttgaatgattgctcccaagaacctgagatccttcaagaatgataa 542
 DB 551 TATGATCATCTATGATGAGATGTTTGAAGAAATCTCTGCAAAAGTGTAGCATGATGAG 610
 QY 543 gacaccaatctgtcttgaatgagagaagtgccacttattggttaataagaagcat 598
 DB 611 GAGAAACATCTATTTGTAATGGAGAGAAATGTCACTTCACTGATGATGATGAT 666

RESULT 2

B67757 680 bp DNA GSS 09-DEC-1997
 LOCUS T24F14TF TAMU Arabidopsis thaliana genomic clone T24F14, DNA
 DEFINITION sequence.
 ACCESSION B67757
 VERSION B67757.1 GI:266511
 KEYWORDS GSS.
 SOURCE
 ORGANISM Arabidopsis thaliana
 thale cress.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 680)
 Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.,
 Berry,K., Granger,D., Sub,E., Wible,C., Adams,M.D. and Venter,
 J.C.

TITLE A BAC End Sequence Database for Identifying Minimal Overlaps in
 JOURNAL Arabidopsis Genomic Sequencing. Update 3
 COMMENT Unpublished (1997)
 Other_GSSs: T24F14TR

Contact: Steve Rounsley
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: rounsley@ligr.org
 Seq primer: M13-21
 Class: BAC ends
 High quality sequence stop: 680.
 Location/Qualifiers
 1. 680
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="T24F14"
 /clone_11b="TAMU"
 /sex="hermaphrodite"
 /note="Vector: BelorAC11; Site_1: HindIII; Site_2: HindIII
 ; Produced by Rod Wing"

BASE COUNT 192 a 143 c 144 g 201 t
 ORIGIN

Query Match 39.4%; Score 298.4; DB 13; Length 680;
 Best Local Similarity 70.1%; Pred. No. 2.3e-71;
 Matches 401; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 186 aataagccactagaagaatcattacccttgccttcttattcagccatagtagaag 245
 DB 1 AATGACGCCACTAGAGAAAGATCATTTCCCTTACTTTCATGACCAATGCTTAGAAGA 60
 QY 246 ttgtcacaactcacacactctgtctcttagaaggtattcttagttctccaaataca 305
 DB 61 TTAGCCATATCCCTATCTGTTCTTAGATGCGTATAGCGGATTTTTCAAATCCCT 120
 QY 306 gtgcacaatctgacagagaagaaacaccttcaacctgccttcggttacccttctat 365
 DB 121 ATCCACCCATATGATCAGAGAAATCAATTCACGTCGTCATGAGACTTTTGTCTTAC 180
 QY 366 agacgtatgctcttggctatgtatgaatgcacctgcaccttcaagaatgataagtgct 425
 DB 181 AAACGATGCTCTTTGGGCTTTGTAAATGACACCCACCTTTTACGCGCTGCATGACTTC 240
 QY 426 atattcctcaactgttgagaataatgtcgaagtttcaatgagttacccttccgttac 485
 DB 241 ATTTTCTCCGATTTGATAGAGAGATGTCAGAGTATCTATGATGATATTTTCTGTAT 300
 QY 486 gggctctcttgaatgattgctcgaagaaccttgatcgaagcttcaagaatgataagac 545
 DB 301 GGCTCTTCTTTCCTCCCTGTTTGTCAATTTTGTGAGGCTACTCAAGATGCGAAGAA 360
 QY 546 accaatctgtcttgaatgaggaagtgccacttattggttaataagaagcatgcttcta 605
 DB 361 ACMAACCTATGCTGTAACGGAAGAAATGTCATCTTATGTTAGAGAAAGAAATGCTTGG 420
 QY 606 ggaacaaattctgaagaaggtatgaagtcgaatgaagtcgaagttatgataagac 665
 DB 421 GGCGACAAATATACAGAAAAAGGATAGAAAGTTGATGGGCAAAATGATGATGATGAT 480
 QY 666 aaaaatgcatacccaagatatacaaggtataagaagtttcttgcattggtggtttc 725
 DB 481 CAGTTGCAACCAACCAAAACCTGTAAGAAATATTAAGAAAGTTTCTCGACATGCTGGCTTC 540
 QY 726 tatagaagttatacaagaattcacaaggt 757
 DB 541 TACAGGAGATCATCATGATTTCTCTAAGTT 572

RESULT 3
 LOCUS A0956470/c 624 bp DNA GSS 28-JAN-2000
 DEFINITION LERAK19TF LERA Arabidopsis thaliana genomic clone LERAK19, DNA

Accession	Sequence
A0956470	sequence.
A0956470	sequence.
A0956470.1	GI:6784171
KEYWORDS	GSI.
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana
REFERENCE	Eukaryotes: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta
AUTHORS	Buell, C.R., Lin, X., Pei, G., Barnstead, M., Bowman, C., Utterbach, T.,
TITLE	Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms
JOURNAL	Unpublished (2000)
COMMENT	Contact: Xinying Lin The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: atligdr.org For additional information, see http://www.tigr.org/tdb/at.at.html Seq primer: TF Class: Shotgun.
FEATURES	Location/Qualifiers
source	1..624 /organism="Arabidopsis thaliana" /strain="Landsberg erecta" /db_xref="taxon:3702" /clone="LERAK19" /clone_11b="LEBA" /note="Organ: Leaf; Vector: pHSO1; Total genomic DNA was sheared to 0.9-1 Kbp before ligation."
BASE COUNT	191 a 142 c 110 g 181 t
ORIGIN	
Query Match	38.2%; Score 288.8; DB 13; Length 624;
Best Local Similarity	69.5%; Pred. No. 9.9e-69;
Matches 392; Conservative	0; Mismatches 172; Indels 0; Gaps
187	ataaagcgaactgagaagatcatctacccttgcctttatgacacatatgtctagaagt 246
624	atgctgtacacagaaagacacatttcccttaccatttatgcatacagatgttgagaggt 565
247	tgctcaaacacacacacactctgcttctctagaaggtatctagtcttccccaataccag 306
564	tagcaattcatatgacttctgcttcttgatgagatctacgaattcttgcagatccgga 505
307	ttgcacaacttgatcagagagaacaacactttcaacttgccttgccttgccttgccttgc 366
504	atcatcccaaatgacacagaaacacacatttcccttgccttgccttgccttgccttgc 445
367	gagctatgaccttggccttgaatgaagcaccctgcacacacacacacacacacacacac 426
444	agagaaagcccttgccttgccttgaatgaagcaccctgcacacacacacacacacacac 385
427	tattcttaactttgtgaagaataatgctgcagaggtttcaatgataagcttccgtttacg 486
384	tcttcacagatgatgacagatgacagatgacagatgacagatgacagatgacagatgac 325
487	ggtcttcttgcagatgac 546
324	gattcatcttgcagatgac 265
547	ccaacttgccttgaatgagagagagatgacacacacacacacacacacacacacacac 606
264	aacatctgattgacagatgacagaaagatgacagatgacagatgacagatgacagatg 205
607	gacataaacttgcagaaagagatgacagatgacagatgacagatgacagatgacagatg 666
204	gacacagagatgacagatgacagatgacagatgacagatgacagatgacagatgacag 145

QY	667	aaagccatacccccacagatatacaagatataagaagtttccttgctacggtgttct	726
Db	144	GTCTTCAAGCGCCGATTAATGTCTAAACAGCTGAGAGCTTCTTGACATGTTGGTTTCT	85
QY	727	atagaagttatacaaatctca 750	
Db	84	ATAGAGATTCATCATCAACACTTTA 61	
RESULT	4		
LOCUS	A0967851/c		
DEFINITION	A0967851 660 bp DNA	GSS	28-JAN-2000
ACCESSION	DERIV29TF LERG Arabidopsis thaliana genomic clone LERIV29, DNA		
VERSION	A0967851		
KEYWORDS	A0967851.1 GI:6795552		
SOURCE	GSS.		
ORGANISM	thale cress.		
	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (bases 1 to 660)		
AUTHORS	Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Uterbach, T.,		
	Feldblyum, T., Liang, F., Creasy, T. and Fraser, C.M.		
	Genomic survey sequencing of Landsberg erecta ecotype of		
	Arabidopsis thaliana and identification of sequence-based		
	polymorphisms		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Xiaoying Lin		
	The Institute for Genomic Research		
	9712 Medical Center Dr., Rockville, MD 20850, USA		
	Tel: 301 838 0200		
	Fax: 301 838 0208		
	Email: atel@ig.ori.org		
	For additional information, see http://www.tigr.org/tdb/at.html		
	Seq primer: TF		
	Class: shotgun.		
FEATURES	Location/Qualifiers		
source	1..660		
	/organism="Arabidopsis thaliana"		
	/strain="Landsberg erecta"		
	/db_xref="taxon:3702"		
	/clone="LERIV29"		
	/clone.lib="LERIV29"		
	/note="Organ: Leaf; Vector: pUC19UK; Total genomic DNA was		
	sheared to 0.4-0.7 kbp before ligation."		
BASE COUNT	195 a 127 c 134 g 204 t		
ORIGIN			
Query Match	35.2%; Score 266.2; DB 13; Length 660;		
Best Local Similarity	67.7%; Pred. No. 1.6e-62;		
Matches 373; Conservative	0; Mismatches 178; Indels	0; Gaps	0.
QY	2	aaaagaggttgtagagcttcctcctgtagtaggtatatactatcatcagtcgtccatagagatg 61	
Db	555	AAAGAGATCTACTTAACCTATTAGATGGGGGTGTAATCTATCCAAATATTCATGACACTTG 496	
QY	62	ggttagcgcggtgtagtagcgttcctcctagaagagagagcatcacgttgctccctaaatgataa 121	
Db	495	GGTATCTGCGACTTATATGTGTGTTCTAAAGAGGAGGTATGACAGTTATTAATAATTCAAA 436	
QY	122	ggaatgaattgatacccgagagagattatcaactgctatagatagatggtgtagatcogattcagaa 181	
Db	435	AGATGAACCTGATACCTAGGACCAATACAGACATAGATGATGTTATGACTATATAGGA 376	
QY	182	actgaataaaccactcagaaagatcaatcccttcgtcccttatacgcacatactgtaga 241	
Db	375	GCTTATGTCACCCACTAGGAAGATCACTTCCCTTACCTTTCAATTGACCAAAATGCTAGA 316	
QY	242	aaggtgtgccaactacacacactctgcttcttctagacaggttatatctagttctcccaat 301	

Matches	360,	Conservative	0;	Mismatches	182;	Indels	1;	Gaps	1;		
OY	213	cccttgcctttatcgcacacatagctagaagaagcttgcacaaactccacacacttgcctt									
Db	9	CCCCCTTCATTATTCAGATCAAAATGTTGGAGGGTTAGCAAAATTCAGAACTGACTACTGTTT									
OY	273	ctgaagcgttatcttgaattctccccaatctacagttgcacatcttgatcagagaanaacc									
Db	69	CTTGATGATATTTACAGGTTTCTTCCAAATCCCGAATTCATCTCGAATGACGAGAGAAAAT									
OY	333	acttccacccttgccttcgtaaccttgccttataagaagatgaccttgccttgccttgc									
Db	129	ACTTTCACTTCCCAATATGATGATTTGCTTATCGAAGAAATGCCCTTTCGCTTGGTAC									
OY	393	gcacactgc-cacacttcaagaatataagatgcatatctctctaacttgccttgccttgc									
Db	189	GCTTCCGCAAACTTTCAGAGATGATGATGCTATCTTTCACGCAATATGATGAGAGATAT									
OY	452	tgtcagagtttccatgtagtaccttccgcttaccggttcctcttgccttgccttgc									
Db	249	CATGAGAGCTTTATTTATGATGATTTTTCAGCTATGAGCAATCATCATTTGAGGACTGTTTGA									
OY	512	caaccttgatcagcgtcttaccagagatgtaaagaaccccaactcttgccttgccttgc									
Db	309	GAATCTTCGCAAGGATTTGGGAGAGATGTGAGAGAGAAACATCTAGTTTGAATTTGGGAAA									
OY	572	gtgcacactttagtgttaaagaagcatcgcttcaagacataaaatttctgaagaagagat									
Db	369	ATGTGACCTTTATGCTGCTAGGATGGGATGGATGTTCTTGCTACATGATTTCTGAGCATGCTAT									
OY	632	tgaagtcgataagcgttaagttgatgatacgcacaaatgcccatacccacagatataa									
Db	429	AGAGGTTATATGAGCCAAAGATCGAAGTCATGACAAACCTTCAGGGCCCTGGAAAGTGA									
OY	692	aggtataagaagattcccttgccttgccttgccttgccttgccttgccttgccttgc									
Db	489	AGCACTGAGAGATTTCCCTAGCTAGCTAGCATGTTTCTTACAAACATTCATCAAAAGATTTTAA									
OY	752	aaa 754									
Db	549	AAA 551									
RESULT 10											
LOCUS	AQ368305	617 bp	DNA	GSS	07-MAR-2000						
DEFINITION	tox00002617r CUG1 Tomato BAC Library Lycopersicon esculentum genomic clone tox00002617r, DNA sequence.										
ACCESSION	AQ368305										
VERSION	AQ368305.1 GI:4221582										
KEYWORDS	GSS.										
SOURCE	tomato.										
ORGANISM	Lycopersicon esculentum										
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Angiosperms; Eudicotyledons; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.										
AUTHORS	1 (bases 1 to 617)										
TITLE	Budiman,M.A., Mao,L., Wood,T.C. and Wing,R.A.										
JOURNAL	A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing										
MEDLINE	Genome Res. 10 (1), 129-136 (2000)										
COMMENT	20113122										
Contact: Wing RA											
Clemson University Genomics Institute											
Clemson University											
100 Jordan Hall, Clemson, SC 29634, USA											
Tel: 864 656 7288											
Fax: 864 656 4293											
Email: rwing@clemson.edu											
Seq primer: GGAACAGCTATGACCATG											
Class: BAC ends											
High quality sequence stop 617.											

FEATURES	source	Location/Qualifiers
		1..617
		/organism="Lycopersicon esculentum"
		/cultivar="Heinz 1706"
		/db_xref="taxon:4081"
		/clone="tox0002617"
		/clone_id="CUGI Tomato BAC Library"
		/tissue_type="Nuclei preparation from leaf"
		/lab_host="E. coli DH10B"
		/note="Vector: pBelOBC 11; Site.1: HindIII; Site.2: HindIII. Tomato is a vegetable crop that ranks second only to potatoes in value and importance. Among plant geneticists and physiologists, tomato represents an ideal diet model beside Arabidopsis and monocot rice to derive genomic information from. To facilitate the genome analysis of tomato, we have constructed a tomato BAC library that is suitable for positional cloning, physical mapping, and genome sequencing. The library contains 129,000 clones and a random sampling of 498 clones indicated an average insert size of 117.5 kb. With 15x haploid and genome equivalents (1C equals 953 Mb) (Arumuganathan and Earle, 1991), the probability to recover any particular sequence is greater than 99%. High stability, large insert and ease in manipulation make BAC libraries the choice for genome sequencing. Pre characterization of a few hundred bases of insert ends will make BAC clones extremely useful for rapid contig assembly (Ventier, Smith, and Hood, 1996). Here we present the construction, characterization of the tomato BAC library, and preliminary analysis of the 1536 tomato BAC end sequences."
BASE COUNT	191 a	150 c 96 g 160 t
ORIGIN		
Query Match	31.1%	Score 235.4; DB 13; Length 617;
Best Local Similarity	62.4%	Pred. No. 4.5e-54;
Matches	385; Conservative	0; Mismatches 231; Indels 1; Gaps 1.
84	cctaagaaggaggcattccggttcgttccctaatagatgaattggtccgcgaagg	143
616	cctaagaaggaggcattccggttcgttccctaatagatgaattggtccgcgaagg	557
144	attatcaatgcatatagatggttgatcgatttcaggaagaactgaataaagccatgga	203
556	cgggtttacgagngaggggtgtgatgattaccacaactaaattcagactgataaaa	497
204	gattcattacccttgccttattatgcacatatagtctagaaggtgtgccaaaccacac	263
496	gacacatttcgatatgcccttcacatgacatgattggtgattgacattccgaaaggtg	437
264	tctctgttctcagaaggttatcttaattctcccaataccagttgcacaaatctgacg	323
436	tacggtttcttcgatgatgatttcgggggttaatacgaattttctatgcacagaaatca	377
324	gagaaacacacttcaactgccttcggttaccttgcattatagacgttatgctttg	383
376	tgaataacacattttactgttcctatgacacatttcggttcgaaagaatgacatttgg	317
384	tatagttaagcaactccacacttcaagaatgatgatgatacttccctaactttgt	443
316	tttttgcaatgaccccccacatttcagacatgatatgctgactttctgcacatggtg	257
444	gaaacattgtcgaagtttcatatgatacttccggttaagggcttcttttgatg	503
256	gagatattcatagaaattttt -atgattgatttttttctgtgttgatttcgtgcgg	198
504	tgcctcagcaacttgatcgaagctctacagaagtgtaagaacacacactgtcgtgaat	563
197	tgcctcagcaacttgatcgaagctctacagaagtgtaagaacacacactgtcgtgaat	138
564	gaggaagatgcaacttctgataagaagatcgtctttagaataaaacttgcga	623
137	tgggaataagtcatttcacgtggaagagcattgtttgggtcatcacaatttcata	78

Oy 182 actgaataagccactagagaagatcattacccttgcctttatcgaccatagtctga 241
 Db 207 GTTAATATGCTGCTTAAGAGAGACCATTTCCCATTTGCTTATTAATGATGCTAGA 266
 Oy 242 aaggttgcacacacacacactctgtcttcttagagagttatcttagtttcccaat 301
 Db 267 AAGATTACAACTATCTTACTATGAGCTTCTTAATGATGATTAAGGATTTTTCAAAT 326
 Oy 302 accagttgacaactctgacagagagaacacacttcccttgccttgcgtatcttgc 361
 Db 327 CTCATATACCAAAATGACCAAGAAACACTTTTACTTGTCTTAATGAGGACCTTGC 386
 Oy 362 ttatagacgtatgccttgccttgccttgccttgccttgccttgccttgccttgc 421
 Db 387 TTACAGCGTATGCTTGTGCTGCTGATATGACCTGCTTACCTTGTGCTGCTGCTG 446
 Oy 422 ggcataatctcacttgccttgccttgccttgccttgccttgccttgccttgccttgc 481
 Db 447 TTCCATTTTCTGCTGATTTGATAGACAGATGCTAGAGATTCATGATGATTTTATGT 506
 Oy 482 tta 484
 Db 507 GTA 509

RESULT 13

A0958534 726 bp DNA GSS 28-JAN-2000
 LOCUS LERAY15TR LERA Arabidopsis thaliana genomic clone LERAY15, DNA
 DEFINITION sequence.
 A0958534
 VERSION A0958534.1 GI:6786235
 KEYWORDS
 SOURCE GSS.
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 726)
 Buell,C.R., Lin,X., Pal,G., Barnstead,M., Bowman,C., Utterbach,T.,
 Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
 Genomic survey sequencing of Landsberg erecta ecotype of
 Arabidopsis thaliana and identification of sequence-based
 polymorphisms
 Unpublished (2000)
 Contact: Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: at@tigr.org
 For additional information, see <http://www.tigr.org/tdb/at.html>
 Seq primer: TR
 Class: shotgun.

FEATURES

Source

1..726
 /organism="Arabidopsis thaliana"
 /strain="Landsberg erecta"
 /db_xref="taxon:3702"
 /clone="LERAY15"
 /clone_lib="LERA"
 /note="Organ: Leaf; Vector: pHOSt; Total genomic DNA was
 sheared to 0.9-1 Kbp before ligation."
 BASE COUNT 230 a 142 c 133 g 221 t
 ORIGIN

Query Match 29.2%; Score 221.2; DB 13; Length 726;
 Best Local Similarity 71.2%; Pred. No. 3; 7e-50;
 Matches 292; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
 Oy 2 aaagaggttgtagactctgtagaggtatcatcatcatcatgcttgctcatagcatg 61

Db 317 AAAGAAATTTTGAAGTATGCTTATGCTGGTGCATCTATACCTATCTCTATGTA 376
 Oy 62 ggtagtcgggtgcatagcgttctctaaagaaggagcatcagctgttccctaatatga 121
 Db 377 GGTTTTCCAGTGATGCTGCTGCTTAAAGGGGATGATGCTGTCTCAAAATGAAAA 436
 Oy 122 ggaagaattgacccgagagatcatcactgagctagagatgtagatcagatccagaa 181
 Db 437 AGATGAATCTGATCCCTTACTATGAACTATTAAGTGTGATGATGATGATGATG 496
 Oy 182 actgaataaagccactagagaagatcattacccttgccttgccttgccttgccttgc 241
 Db 497 GTTGATGCTGCTATGAGAAAGATCATTTTCTTACATTCATGATGATGATGATG 556
 Oy 242 aaggttgcacacacacacactctgtcttcttagaaggttatcttagtttcccaat 301
 Db 557 ACATTTACTTATCATCATATTTATTTGCTTGTGATGATGATGATGATGATGATG 616
 Oy 302 accagttgacaactctgacagagagaacacacttccacttgccttgccttgccttgc 361
 Db 617 ACCAATTCACCTTATATGATCAAGAAAAACACCTTTACCGTGTCTTATGGA 676
 Oy 362 ttatagacgtatgccttgccttgccttgccttgccttgccttgccttgccttgc 411
 Db 677 CTATAGAGAAAGCATTTGTTATGCAATGCTCTGCAACATTTTCA 726

RESULT 14

A0967852 510 bp DNA GSS 28-JAN-2000
 LOCUS LERIV29TRB LERG Arabidopsis thaliana genomic clone LERIV29, DNA
 DEFINITION sequence.
 A0967852
 VERSION A0967852.1 GI:6795553
 KEYWORDS
 SOURCE GSS.
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 510)
 Buell,C.R., Lin,X., Pal,G., Barnstead,M., Bowman,C., Utterbach,T.,
 Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
 Genomic survey sequencing of Landsberg erecta ecotype of
 Arabidopsis thaliana and identification of sequence-based
 polymorphisms
 Unpublished (2000)
 Contact: Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: at@tigr.org
 For additional information, see <http://www.tigr.org/tdb/at.html>
 Seq primer: TR
 Class: shotgun.

FEATURES

Source

1..510
 /organism="Arabidopsis thaliana"
 /strain="Landsberg erecta"
 /db_xref="taxon:3702"
 /clone="LERIV29"
 /clone_lib="LERG"
 /note="Organ: Leaf; Vector: pUC19TK; Total genomic DNA was
 sheared to 0.4-0.7 Kbp before ligation."
 BASE COUNT 172 a 104 c 91 g 142 t
 ORIGIN

Query Match 28.4%; Score 214.8; DB 13; Length 510;
 Best Local Similarity 69.5%; Pred. No. 2e-48;
 Matches 291; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
 Oy 2 aaagaggttgtagactctgtagaggtatcatcatcatcatgcttgctcatagcatg 61

7.

Search completed: April 1, 2002, 22:13:19
Job time: 7761 sec

•
•
•
•

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2002, 21:29:33 ; Search time 79.08 Seconds
(Without alignments)
235.109 Million cell updates/sec

Title: US-09-586-106-63
Perfect score: 1347
Sequence: 1 KEVVKLDGEGYIYVHNSDM.....IRSLGCGEYRRIKFTK 251

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /SID2/gcgdata/geneeq/AA1980.DAT:*
2: /SID2/gcgdata/geneeq/AA1981.DAT:*
3: /SID2/gcgdata/geneeq/AA1982.DAT:*
4: /SID2/gcgdata/geneeq/AA1983.DAT:*
5: /SID2/gcgdata/geneeq/AA1984.DAT:*
6: /SID2/gcgdata/geneeq/AA1985.DAT:*
7: /SID2/gcgdata/geneeq/AA1986.DAT:*
8: /SID2/gcgdata/geneeq/AA1987.DAT:*
9: /SID2/gcgdata/geneeq/AA1988.DAT:*
10: /SID2/gcgdata/geneeq/AA1989.DAT:*
11: /SID2/gcgdata/geneeq/AA1990.DAT:*
12: /SID2/gcgdata/geneeq/AA1991.DAT:*
13: /SID2/gcgdata/geneeq/AA1992.DAT:*
14: /SID2/gcgdata/geneeq/AA1993.DAT:*
15: /SID2/gcgdata/geneeq/AA1994.DAT:*
16: /SID2/gcgdata/geneeq/AA1995.DAT:*
17: /SID2/gcgdata/geneeq/AA1996.DAT:*
18: /SID2/gcgdata/geneeq/AA1997.DAT:*
19: /SID2/gcgdata/geneeq/AA1998.DAT:*
20: /SID2/gcgdata/geneeq/AA1999.DAT:*
21: /SID2/gcgdata/geneeq/AA2000.DAT:*
22: /SID2/gcgdata/geneeq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	999	74.2	1801	21	AAV32434
2	790	58.6	200	21	AAV32431
3	425	31.6	1217	20	AAV14565
4	225.5	16.7	916	21	AAV03130
5	225.5	16.7	916	21	AAV03130
6	219.5	16.3	1871	21	AAV57165
7	219.5	16.3	1871	21	AAV57165
8	217.5	16.1	1871	20	AAV01078
9	197.5	14.7	1156	21	AAV01078
10	195.5	14.5	401	18	AAV32092
11	195.5	14.5	401	22	AAV3283

12	190.5	14.1	1194	18	AAV39272
13	189.5	14.1	1145	18	AAV32097
14	189.5	14.1	1145	22	AAV3286
15	174.5	13.0	1784	11	AAV05898
16	174.5	13.0	1784	12	AAV94427
17	169	12.5	740	17	AAV1516
18	167.5	12.4	1203	14	AAV81572
19	165.5	12.3	524	20	AAV41681
20	164.5	12.2	524	21	AAV03348
21	163.5	12.1	1196	16	AAV75189
22	162.5	12.1	672	21	AAV12457
23	158.5	11.8	429	19	AAV68403
24	158.5	11.8	498	16	AAV73936
25	158.5	11.8	498	21	AAV03349
26	158.5	11.8	504	16	AAV27322
27	158.5	11.8	504	18	AAV03350
28	158.5	11.8	504	21	AAV03163
29	158.5	11.8	1079	20	AAV12994
30	158.5	11.8	1199	21	AAV17947
31	158.5	11.8	1224	20	AAV17947
32	158.5	11.8	1737	21	AAV10044
33	151.5	11.2	1197	22	AAV49468
34	134	9.9	1053	11	AAV05614
35	130	9.7	768	19	AAV71062
36	126	9.4	372	22	AAV81066
37	124.5	9.2	1014	9	AAV80810
38	124.5	9.2	1027	13	AAV81773
39	124.5	9.2	1036	13	AAV20599
40	124.5	9.2	1036	21	AAV51978
41	123.5	9.2	1055	18	AAV13055
42	123.5	9.2	3080	10	AAV93285
43	123.5	9.2	3210	9	AAV81771
44	123	9.1	796	18	AAV36024
45	123	9.1	796	19	AAV71080

ALIGNMENTS

RESULT 1	
AAV32434	
ID	AAV32434 standard; Protein: 1801 AA.
XX	
AC	AAV32434;
XX	
DT	27-MAR-2000 (first entry)
XX	
DE	Plant generic retroelement.
XX	
KW	Retroelement; retrovirus; transgenic plant; gene transfer;
KW	soybean; pea; calypso; athila; cyclops.
XX	
OS	Glycine max.
OS	Arabidopsis thaliana.
OS	Pisum sativum.
XX	
PN	WO9960842-A2.
XX	
PD	02-DEC-1999.
XX	
PF	28-MAY-1999; 99WO-US11858.
XX	
PR	29-MAY-1998; 98US-0087125.
XX	
PR	28-MAY-1999; 99US-0087125.
XX	
PA	(WRIGHT) WRIGHT D A.
XX	
PA	(VOYT) VOYTAS D F.
XX	
PI	Wright DA, Voytas DF.
XX	
DR	WPI: 2000-105586/09.
XX	
DR	N-PSDB; AA234261.
XX	

Porcine retrovirus
Minature swine re
Retroviral protein
Gene product of fl
Pely F6a provirus
Soybean chlorotic
Mus dunni endogeno
Reverse transcript
Osteoinductive ret
Moloney murine leu
Fely-A pol protein
Novel reverse tran
M-MLV reverse tran
Novel reverse tran
Reverse transcript
M-MLV reverse tran
MLV reverse trans
MLV reverse trans
MOMLV pol gene pro
MLV gag-pol prote
Canine retrovirus
BIV pol gene produ
Multiple sclerosis
HIV protease and r
Sequence of pol pr
Sequence encoded b
ROD HIV-2 polymera
HIV-2 ROD isolate
HIV-2 provirus-enc
Sequence of clone
Deduced sequence e
MSRV-1 pol protein
Multiple sclerosis

1

KW Genome; retroelement; retrovirus; ZAM; gag; pol; env; LTR; vector;
KW long terminal repeat; gene therapy.
XY

XX	Drosophila melanogaster.
XX	
PN	FR2772045-A1.
XX	
PD	11-JUN-1999.
XX	
PF	10-DEC-1997; 97FR-0015655.
XX	
PR	10-DEC-1997; 97FR-0015655.
XX	
PA	(UYAU-) UNIV AUVERGNE.
XX	
PI	Leblanc P, Vaury C;
XX	
DR	WPI; 1999-359998/31.
DR	N-PSDB; AAX83445.
XX	
PT	New Drosophila ZAM retroelement nucleic acid - useful in gene transfer or gene therapy
XX	
PS	Disclosure; Fig 1; 55pp; French.
XX	
CC	This sequence represents the pol protein from a novel retroelement ZAM found in Drosophila melanogaster. Similarly to other retroelements and retroviruses, ZAM contains 3 genes (gag, pol and env) and long terminal repeats (LTR's) at either end of the genome. Vectors containing the ZAM nucleic acid sequence can be used for gene transfer or gene therapy.
CC	
XX	
SO	Sequence 1217 AA;

```

Query Match 425; DB 20; Length 1217;
Best Local Similarity 37.5%; Pred. No. 8, 9e-37;
Matches 94; Conservative 46; Mismatches 97; Indels 14; Gaps 3

Qy      1 KEVYKLLDEGIIVYHAHSDVSPVSHVPPKKGGITVVPNDKDELLPQRITITGYRWIDFRK 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      334 kqlesmlldqgll-rsshpwsaprvwvppk--ldgtgnrk-----wrlvidgtrk 379

Qy      61 LNKATRKDHPLEPFIDHMLERLSKLTHTFCFLDGVSSFSQIPVANSODGKTPFCPTFA 120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      380 lndkltisdyrplnldldlsigkayfslcltsgthgiempkxlakafavveghye 439

Qy      121 YRDMPGLCAVAPPTFFOCMAIFSNFECENIVEVDFSVYSSDFDCLSLNDVLRQK 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      440 ftmrfglnkapatftqvmvdsavgldngtclcflyldllflspqlkhlldikmvfeclr 499

Qy      181 DTNLVLNGEKCHPFWNEGIVLGHKISERGIIEVDKAKYDAIDKMPYPRIDKIGRSFLGHG 240
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      500 aanfklqpskseflrkeleflghivtgdgyknpnklsaklkkfcpennraiksflgllg 559

Qy      241 FYRRFIKDFPK 251
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      560 yrrkfrldfar 570

RESULT 4
AAB03130
ID AAB03130 standard; Protein: 916 AA.
AC AAB03130;
XX
DT 10-OCT-2000 (first entry)
XX
DE Gag-pol readthrough protein encoded by C. albicans retrotransposon 15.
XX
KW Retrotransposon; pCal; TCA2; Ty1; copia; long terminal repeat; LTR;
KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
KW stop codon suppression; gene delivery; gene therapy vector;
KW genetic vaccine composition; immunogenic; transgenic animal.
XX

```

OS	Candida albicans.
xx	
PN	MO2000026397-A1.
xx	
PD	11-MAY-2000.
xx	
PF	01-NOV-1999; 99MO-NZ00179.
xx	
PR	30-OCT-1998; 98CA-2249046.
PR	30-OCT-1998; 98US-0106342.
xx	
PA	(JANC) JANSSEN PHARM NV.
xx	
PI	Luyten WHML, De Backer MD, Neijssen BJM, Poulter RTM;
xx	
DR	WPI: 2000-365640/31.
DR	N-PSDB; AAA57953.
xx	
PT	Novel retrotransposon expression vectors useful for expressing an antigen, epitope or therapeutic agent, or detecting genes or the presence of Candida in a sample -

Example 12; Fig 33; 204pp; English.

The invention relates to novel retrotransposons from the yeast *Candida albicans* which have a copy number of 40-150, preferably 50-100 copies per genome. In particular, the invention relates to the novel *C. albicans* TY1/copia retrotransposon *pcal* (AA57920), and to the integrated form of this retrotransposon, designated Tca2, and to the novel *C. albicans* retrotransposons 1-28. *pcal* was initially isolated from *C. albicans* HOG1042 and has a copy number of 50-100 copies per cell. It comprises identical 280 bp long terminal repeats (LTRs) and two open reading frames (ORFs). The first ORF encodes a gag (group antigen) protein, and the second ORF encodes a polyprotein (pol) consisting of an aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and pol ORFs of *pcal* are in the same reading frame, separated only by a termination codon (TGA). Translation of the pol ORF occurs through the occasional readthrough suppression of a pseudoknot within the gag-pol mRNA. The retrotransposons of the invention can be used as vectors for in vitro or in vivo transformation and expression. They can thus be used for the delivery and expression of a therapeutic, immunological or immunogenic molecule (e.g., an antigen) and may also be used for eliciting an immunological response in a host organism. They are therefore useful in genetic vaccine compositions and for gene therapy, particularly where the use of retroviral vectors is unsafe or undesirable. Additionally, the retrotransposons may be used to generate transgenic animals, to detect the presence of *Candida* in a sample, to detect and disrupt genes, and to assign functions to nucleotide sequences. Sequences AAB03129-B03133 represent proteins or protein fragments encoded by the novel *C. albicans* retrotransposons of the invention.

	Query Match	Similarity	Best Local Matches	Score	DB	Length	gaps
				16.7%	225.5	916	
				24.2%	No. 2.3e-15		
				48;	Mismatches 112;	Indels 31;	Gaps 6;
Oy	7	LDEGIIVHVASDWPVSFVHSVPKGGITVVPNDRDELIPORLITYGRMYIDFRKLANKA-	-65				
Dd	273	leagvvlppapldwlpripfirktna-----nq-----sstklavdlrlrlnkvev	317				
Oy	66	RKHHYLPFLPDHMLERSKLTTHCFLDGVSSFSQIIPAASDOEKTFPTCPGFAYRRMP	125				
Dd	318	rmytyrpdcctkd-lisslscshysalclknatyaivslnksikyfyistsegunycyttlr	376				
Oy	126	FGLCNAPATFORCMMAIFSNFCENIVE----VEMDDSFVGSSSFDCLSLNDRVLQR	179				
Dd	377	fgaainsp-----tftfnfvrgilegipcfifymddallihcktlhdmsllrrimekl	428				
Oy	180	KDTNLAVLNKGECCHPMVEAGIYLVGHKISRREIEVDKKAKVIDAMKPYPDTIKIGRSP	239				

CC for producing plants having unique signature sequences or other marker
CC sequences which can be used to identify proprietary lines or varieties,
CC and transgenic plants for commercial and research purposes. Transgenic
CC plants may be created for use in traditional agriculture and for use in
CC the manufacture of proteins or other compounds, which are extracted or
CC purified from plant parts or seeds. They may also be used in commercial
CC breeding programmes, including in the creation of new mutant plants
CC through insertional mutagenesis.

Sequence 1871 AA;
SQ

Query Match	16.3%;	Score 219.5;	DB 21;	Length 1871;
Best Local Similarity	30.2%;	Pred. No. 2.8e-14;		
Matches	61;	Conservative	37;	Mismatches 99;
			Indels	5;
			Gaps	3;

[illegible]

RESULT 7
AAV57165
ID AAV57165 standard; Protein; 1871 AA.

DT 11-FEB-2000 (first entry)

DE Amino acid sequeunce of sugarcane bacilliform virus gene.

KW Sugarcane bacilliform virus; SCBV; promoter; transgenic plant;

KW monocot plant; dicot plant.

OS Sugarcane bacilliform virus.

PN US5994123-A.

PD 30-NOV-1999.

PF 09-AUG-1996; 96US-0694869.

PR 09-AUG-1996; 96US-0694869.

PA (MINU) UNIV MINNESOTA.

PI Lockhart B, Torbert K, Olszewski N, Tzafir I, Somers DA;

DR WPI; 2000-038263/03

XX
XX
XX

proteins and RNA transcripts in transgenic plants -

aa Disclosure; Fig 1A-B; 34pp; English.
ps

The invention provides sugarcane bacilliform virus (SCBV) promoters selected from sequences shown in AA23511-513. The promoter is used to express proteins and RNA transcripts in transgenic plants in order to improve their characteristics. The promoter confers constitutively high levels of expression of operably linked preselected DNA sequences in

CC both monocot and dicot plants, plant parts or plant cells.
CC The present sequence represents the sugarcane bacilliform virus amino
CC acid sequence.

acid sequence

Sequence	1871	AA
SQ		

Query Match	16.3%;	Score 219.5;	DB 21;	Length 1877;
Best Local Similarity	30.2%;	Pred. No. 2.8e-14;		
Matches	61;	Conservative	37;	Mismatches 99;
			Indels	5;
			Gaps	3

QY	53	RMVDEPKLNKATKPKKHDPPELFDNHLERISKCTHHCCEFDGYSFSSQJIPVAKOSODEKTEF	112
Db	1316	rlvlnvkrfhnctvpdpsjtpgjnallkvnarekiskfdlksghnyamdeesjpltaf	1375
QY	113	TCPGFTFVARRMRPGLCNABATFQRCMMALFNSFCENIVEFMDESVYSGSFDDCSNLS	172
Db	1376	sayneljyewlvmprgjlknabatfgtrkmdqcfgr-tesgfavlylddl1vtsedeeghaeh1	1434
QY	173	DRVLQRCDDTNVLNNGEKCFHVNNEGIVYGHKISESGIVD---AKVDAIDKMPYPRDI	229
Db	1435	wkmlqickrtnglllspskykrygvkvdflgstlgdqlqavgenllkkaiefdekltk-	1493
QY	230	KGIRSEFLGGEVFRFLEKDETK	251
Db	1494	eglkswlatltnyarnhikdmgk	1515

RESULT	8
AAV01078	
ID	AAV01078 standard; Protein; 1871 AA

AC AAY01078;

DT 08-JUN-1999 (first entry)

DE Sugarcane bacilliform virus protein sequence

KW Sugarcane bacilliform virus; SCBV promoter; transgenic plant,

KW nutritional content; protein manufacture.

OS Sugarcane bacilliform virus.

PN W09909190-A1

PD 25-FEB-1999.

13-AUG-1997; 97WO-IB01338

PR 13-AUG-1997; 97WO-IB01338

AA
PA (MINN) UNIV MINNESOTA

Lockhart B, Olszewski N, Somers DA, Torbert K, Tzafir I;

AA
DR WPI: 1999-181047/15.

DR N-FSDB, AAAA/XXX.
XX

PT promoter - useful for generating fertile transgenic plants with

XX
beneficial agromomic characters

PS Disclosure; Fig 1, 0/PP, Eng+ren
XX

CC This sequence represents a sugarcane bacilliform virus protein.
CC The invention relates to a preselected DNA segment, consisting of
CC a sugarcane bacilliform virus (SCBV) promoter, (including a biologically
CC active subunit). Fertile transgenic plants produced by transformation
CC with the DNA sequence are useful for conventional plant breeding official
CC programs. Other transgenic plants are useful for general deficit, pest
CC abiotic, herbicide resistance, including resistance to water deficit, pest
CC resistance, herbicide resistance, increased yield, improved nutritional

CC content, and improved processing characteristics. They are also useful
 CC for commercial protein manufacture. The SCBV promoter is a strong,
 CC constitutive promoter in many cells, enabling the expression of genes in
 CC fertile transgenic monocot or dicot plants, which allows genetic
 CC engineering of improved plant characteristics.

XX Sequence 1871 AA;

Query Match 16.1%; Score 217.5; DB 20; Length 1871;
 Best Local Similarity 29.7%; Pred. No. 4.7e-14;
 Matches 60; Conservative 38; Mismatches 99; Indels 5; Gaps 3;

QY 53 RMVIDFRKLNATKRDHYPLFPIDHMLERLSKLTFFCLDYSFSQIPVAOSDQKPTTF 112
 Db 1316 RLVNFKRLINDLTPDQSLGPIGALLKVARAKIFSKFDLKSFGHYAMDEESIPLAF 1375
 QY 113 TPCPTGATVRRMPFGCLNAPATFORCMAAIFSNFCENIVEVPMDFSVYSSFPDCLSNL 172
 Db 1376 SAYNELYEWLMPFGIKNAPALFGRKMDQCFRG-LEGFIAYIDDLVFSDEEGHAEHL 1434
 QY 173 DRVLQRCRDTLVNNGECHEFMVNEGIVLGHKISERGIIEVDK---AKYDAIDKMPYPTDI 229
 Db 1435 WKMLQIKRNGIILSPSKYKIGVKVDFLGSTIGANGIAVGENHLIKKIAEFDDEKLTK- 1493
 QY 230 KGRSFLGSGGCFRRRTKDFTK 251
 Db 1494 EGIKSWIATLNGYARNHLKMDGK 1515

RESULT 9

AA10104
 ID AAB10104 standard; Protein; 1156 AA.

AC AAB10104;
 DT 10-NOV-2000 (first entry)

DE Feline foamy virus pol protein.

KW Retroviral vector; anti-FIV; anti-HIV; vaccine; vaccination; cat;
 KW feline immunodeficiency virus; human immunodeficiency virus; FIV; HIV;
 KW pol protein.

OS Feline foamy virus.

PN DE19858441-A1.

PD 21-JUN-2000.

PF 17-DEC-1998; 98DE-1058441.

PR 17-DEC-1998; 98DE-1058441.

PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.

PI Fluegel R, Loechelt M, Flower R, Winkler I;

DR WPI: 2000-432420/38.

DR N-PSDB: AAA40381.

PT Feline foamyvirus retroviral vector for expression of foreign DNA in
 PT animals, especially for vaccination of cats and humans against feline
 PT or human immunodeficiency virus (FIV) or (HIV), respectively -

PS Example 1; Fig 2; 11pp; German.

CC This invention describes a novel retroviral vector to introduce a
 CC desired, expressible DNA into an animal cell, where the retroviral
 CC vector comprises a first DNA sequence, which corresponds to a part of
 CC the reverse transcript of a feline foamy virus (FeV), and a second DNA
 CC sequence permitting propagation in bacteria. The products of the
 CC invention have anti-FIV and anti-HIV activity and can be used in

CC vaccines. The retroviral vector can be used to produce a medicament for
 CC vaccination, in particular for vaccination of cats against feline
 CC immunodeficiency virus (FIV) or humans against human immunodeficiency
 CC virus (HIV). This sequence represents the feline foamy virus pol protein
 CC which is described in the method of the invention.

XX Sequence 1156 AA;

Query Match 14.7%; Score 197.5; DB 21; Length 1156;
 Best Local Similarity 25.9%; Pred. No. 3.5e-12;
 Matches 65; Conservative 47; Mismatches 108; Indels 31; Gaps 6;

QY 6 LLDDEITIVHNSDMVSPVSHVSKGGLTVVNDKDELIPQRIINGYRWVIDFRKLNAT 65
 Db 182 ILKQVLIQ-KESTLMTVPYVPK-----PNGF-----WRMVIDYRAVNKVT 222
 QY 66 RKDHYPLFPIDH-----MLERLSKLTFFCLDYSFSQIPVAOSDQKPTTFCEGTF 120
 Db 223 -----PLIAYQNGHSYGLIGSLFGRYKTCIDLSNGFWAHPLVPEDYWTAFITWYKGYC 277
 QY 121 YRRMPFGLCNAPATFORCMAAIFSNFCENIVEVPMDFSVYSSFPDCLSNLDRVLQRC 180
 Db 278 WTVLPQGLNSPGLFTGVDVGLIGIPN--VEYVDVYISHSEKENLEYIDLIFNLK 335
 QY 181 DTNLVNGECHEFMVNEGIVLGHKISERGIIEVDKAVDAIDKMPYPTDIKIRSFGLHGG 240
 Db 336 EAGYIISIKSNLSANSLVDLIGITLNEGRGLTDFIKENLITAPLTKLQSLIGL 395
 QY 241 FYRRFIRKDFTK 251
 Db 396 FARNFIDPTE 406

RESULT 10

AAW32092
 ID AAW32092 standard; Protein; 401 AA.

AC AAW32092;

DT 09-FEB-1998 (first entry)

DE Porcine retrovirus partial POL protein from defective virus.

KW Retrovirus; porcine; POL protein; xenotransplantation; infectious;
 KW provirus; organ transplant; donor; activated virus; PCR.

OS Porcine retrovirus.

FN Key Location/Qualifiers

FT Protein 1..401

FT /label= POL protein

FT /note="fragment containing putative partial POL viral

FT protein"

FN WO9721836-A1.

PD 19-JUN-1997.

PF 13-DEC-1996; 96WO-US19680.

PR 14-DEC-1995; 95US-0572645.

PA (GEHO) GEN HOSPITAL CORP.

PI Fishman JA;

DR WPI: 1997-332804/30.

DR N-PSDB: AAT74883.

PT New nucleic acid from porcine retroviruses - used for detecting
 PT viruses in transplant or other tissue and for assessing risk of
 PT transmitting infection to graft recipient

XX Claim 16; Fig 2; 128bp; English.

PS This sequence represents a putative partial viral POL protein from a
 CC defective purified swine retrovirus found in PK-15 cells. There are a
 CC few in frame stop codons and apparent frame shifts in the given cDNA
 CC coding sequence which alter features of the translation
 CC (see AAM32093-W32095 for other parts of the sequence). This sequence and
 CC PCR fragments generated from the sequence (see AAT74812-T74882) can be
 CC used to screen organs for porcine retroviruses prior to
 CC xenotransplantation. Transplantation can increase the likelihood of
 CC retroviral activation if intact and infectious proviruses are present.
 CC The porcine retroviral sequence can be used to generate probes to
 CC determine the level (e.g. copy number) of intact (i.e. potentially
 CC replicating) porcine provirus sequences in a strain of xenograft
 CC transplantation donors. It can be used to detect mutations, genetic
 CC lesions or viral recombinants and also to determine the histological
 CC localization of activated retrovirus. Using Polymerase Chain Reaction DNA
 CC Quantitation (PQD) on blood mononuclear cells, infectivity titration and
 CC susceptibility testing can be performed. Ultimately animal donors without
 CC intact porcine retroviral sequences or a lower copy number of viral
 CC elements could be selected.

XX SQ Sequence 401 AA;

Query Match 14.5%; Score 195.5; DB 18; Length 401;
 Best Local Similarity 26.0%; Pred. No. 1.3e-12;
 Matches 67; Conservative 42; Mismatches 120; Indels 29; Gaps 6;

QY 3 VVKLDEGIIYVAHSDWSPVHSVPKKGITVVPNDKDELIPQRIITGYRMVIDPRKLN 62
 DB 147 Vqrlldqglllvpr-qspwnrlpllvprkpg-----tnd-----yrrpqdlrevn 188
 QY 63 KATRKMDYPLPFTIDHMLERL-SKLTNFCPLDGYSSFSQIPVAOSDDEKTF-----TCP 115
 DB 189 kvrgdlhprtpnpnyllsalprernwylvldlkdaefclrlhpsqplffewrptgtgr 248
 QY 116 FGTFAVRMPFGICNAPATQRCMAIFSNFCENIVEY----FMDDFSVYGSFDDCLSN 171
 DB 249 tggltwrlrpgfknspflfdeahlrldanfrlqhpqvlllyvddlllagatkqdcleg 308
 QY 172 LDRVLOKCKDNLVNEKCHFMWNEGIVLGHKISEKGEVDKAKYDAIDKMPYPTDIK 231
 DB 309 tkaillesldlgyrasakkaqicrrevlylgyrlrgqgwltearkktvwqipaptakg 368
 QY 232 IRSFLGHGFGYRRFIKDF 249
 DB 369 vreflgtagfcrllwipgf 386

RESULT 11

AAB73283
 ID AAB73283 standard; Protein; 401 AA.

XX AAB73283;

XX 23-MAY-2001 (first entry)

XX Defective retroviral genome protein #2 isolated from PK-15 cell line.

XX Retrovirus; graft transplantation; xenotransplantation; PK-15 cell line.

XX Unidentified.

XX US6190861-B1.

XX 20-FEB-2001.

XX 13-DEC-1996; 96US-0766528.

XX 14-DEC-1995; 95US-0572645.

XX PR

PA (GEHO) GEN HOSPITAL CORP.

XX Fishman JA;

XX WPI; 2001-256211/26.

XX N-PSDB; AAF77726.

XX Assessing risk of endogenous retroviruses in clinical practice and in
 PT xenotransplantation, comprises using probe sequences derived from swine
 PT or miniature swine retroviral genome

XX Disclosure; Fig 2; 127bp; English.

XX The present invention relates to a method for screening a cell or tissue
 CC for the presence or expression of a retrovirus (RV), comprising
 CC contacting a target nucleic acid from the cell or tissue with a second
 CC nucleic acid from the present invention (e.g. AAF77726 or a fragment
 CC thereof). The method is useful for RV detection and to assess graft
 CC transplantation risk. Screening of animals allows the elimination of
 CC donors with active replication of known viruses. Inactive proviruses can
 CC be detected and inactivated, allowing identification and elimination of
 CC potential human pathogens derived from swine in a manner not possible in
 CC the outbred human organ donor population and is important to the
 CC development of human xenotransplantation.

XX SQ Sequence 401 AA;

Query Match 14.5%; Score 195.5; DB 22; Length 401;
 Best Local Similarity 26.0%; Pred. No. 1.3e-12;
 Matches 67; Conservative 42; Mismatches 120; Indels 29; Gaps 6;

QY 3 VVKLDEGIIYVAHSDWSPVHSVPKKGITVVPNDKDELIPQRIITGYRMVIDPRKLN 62
 DB 147 Vqrlldqglllvpr-qspwnrlpllvprkpg-----tnd-----yrrpqdlrevn 188
 QY 63 KATRKMDYPLPFTIDHMLERL-SKLTNFCPLDGYSSFSQIPVAOSDDEKTF-----TCP 115
 DB 189 kvrgdlhprtpnpnyllsalprernwylvldlkdaefclrlhpsqplffewrptgtgr 248
 QY 116 FGTFAVRMPFGICNAPATQRCMAIFSNFCENIVEY----FMDDFSVYGSFDDCLSN 171
 DB 249 tggltwrlrpgfknspflfdeahlrldanfrlqhpqvlllyvddlllagatkqdcleg 308
 QY 172 LDRVLOKCKDNLVNEKCHFMWNEGIVLGHKISEKGEVDKAKYDAIDKMPYPTDIK 231
 DB 309 tkaillesldlgyrasakkaqicrrevlylgyrlrgqgwltearkktvwqipaptakg 368
 QY 232 IRSFLGHGFGYRRFIKDF 249
 DB 369 vreflgtagfcrllwipgf 386

RESULT 12

AAW39272
 ID AAW39272 standard; Protein; 1194 AA.

XX AAW39272;

XX 19-MAY-1998 (first entry)

XX Porcine retrovirus POL protein.

XX Porcine retrovirus; POL protein; ENV protein; GAG protein;

XX vaccine; diagnosis; xenotransplantation; prophylactic; therapeutic.

XX Porcine retrovirus.

XX WO9740167-A1.

XX 30-OCT-1997.

XX 18-APR-1997; 97WO-GB01087.

XX PF

XX 10-FEB-1997; 97GB-0002668.
 PR 19-APR-1996; 96GB-0008164.
 XX (IMUT-) IMUTRAN LTD.
 PA (COONE-) O-ONE BIOTECH LTD.
 XX Galbraith DN, Haworth C, Lees GM, Smith KT;
 XX WPI: 1997-535851/49.
 DR N-PSDB: AAV09700.
 XX
 PT Polynucleotide encoding porcine retrovirus expression product -
 PT useful to develop products for use in vaccines, diagnosis and
 PT xeno-transplantation
 PS
 XX Claim 6; Fig 3; 69pp; English.
 CC This sequence represents the porcine retrovirus (PoEV) polymerase (POL)
 CC protein. This protein and other porcine retroviral proteins e.g. the
 CC virion core (GAG) and envelope (ENV) proteins can be used to develop
 CC viral vaccines, antisense nucleic acids, ribozymes and other antiviral
 CC agents. They can also be used in xeno-transplantation technology and as
 CC diagnostic tools.
 XX
 SO Sequence 1194 AA;

Query March 14.1%; Score 190.5; DB 18; Length 1194;
 Best Local Similarity 25.6%; Pred. No. 2.1e-11;
 Matches 66; Conservative 42; Mismatches 121; Indels 29; Gaps 6;

QY 3 VKLLDEGIIVAHSDWSPVHSYPRKGGITVVPNDKDELIPQRIITGYRMVIDFRKLN 62
 Db 196 Vgrllggllvpr-qspntrllpvrkpg-----tnd-----yrvpqlirevn 237
 QY 63 KATRKDHPLPFIIDHMERL-SKLTHFCFLDGYSSFSQIPVAQSDQKTTF-----TCP 115
 Db 238 krvgdlhptvnpynllsalppernwylvldkdaifclrlhptsqplfafewrdpgtgr 297
 QY 116 FGFAVRRMPFGLCNAPATFORCMMAIFSNFCENIVEV----FMDFSVYSSPDDCLSN 171
 Db 298 tggltwtclpggfknsplfdeahrdlanfrlqpgvcllgyvddlllagatkqdcleg 357
 QY 172 LDRVLQRCQDTNLVNGECKHFWMNEGIVLGHKISERGIEVDKAKVDAIDKMPYPTDIKG 231
 Db 358 tkallllesdlgyrasakkaqicrrevlylgsylrsggrwltearkkkvvgjapptakg 417
 QY 232 IRSFLGHGFGYRRFIKDF 249
 Db 418 vreflgtagfcrllwipgf 435

RESULT 13
 AAMW32097
 ID AAMW32097 standard; Protein; 1145 AA.
 XX
 AC AAMW32097;
 XX
 DT 09-FEB-1998 (first entry)
 XX
 DE Miniature swine retrovirus POL protein.
 XX
 KM Retrovirus; porcine; POL protein; xenotransplantation; infectious;
 KM provirus; organ transplant; donor; activated virus; PCR.
 XX
 OS Porcine retrovirus.
 XX
 FH Key Location/Qualifiers
 FT Protein 1..1145
 XX /label= POL_protein
 XX
 PN MO9721836-A1.

XX 19-JUN-1997.
 PD
 XX 13-DEC-1996; 96MO-US19680.
 PF
 XX 14-DEC-1995; 95US-0572645.
 PR
 XX (GEHO) GEN HOSPITAL CORP.
 PA
 XX Fishman JA;
 PI
 XX WPI: 1997-332804/30.
 DR N-PSDB: AAT74884.
 XX
 PT New nucleic acid from porcine retroviruses - used for detecting
 PT viruses in transplant or other tissue and for assessing risk of
 PT transmitting infection to graft recipient
 PS
 XX Claim 22; Fig 3; 128pp; English.

This is a porcine retrovirus from miniature swine containing the coding region for a putative viral POL protein. This sequence and PCR fragments generated from the sequence (see AAT74812-T74882) could be used to screen organs for porcine retroviruses prior to xenotransplantation. CC Transplantation can increase the likelihood of retroviral activation if CC intact and infectious proviruses are present. The porcine retroviral CC sequence can be used to generate probes to determine the level (e.g. CC copy number) of intact (i.e. potentially replicating) porcine provirus CC sequences in a strain of xenograft transplantation donors. It can be CC used to detect mutations, genetic lesions or viral recombinants and CC also to determine the histological localisation of activated retrovirus. CC using Polymerase Chain Reaction DNA Quantitation (PDQ) on blood CC mononuclear cells, infectivity titration and susceptibility testing can CC be performed. Ultimately animal donors without intact porcine retroviral sequences or a lower copy number of viral elements could be selected.

Sequence 1145 AA;

Query March 14.1%; Score 189.5; DB 18; Length 1145;
 Best Local Similarity 25.6%; Pred. No. 2.5e-11;
 Matches 66; Conservative 44; Mismatches 119; Indels 29; Gaps 6;

QY 3 VKLLDEGIIVAHSDWSPVHSYPRKGGITVVPNDKDELIPQRIITGYRMVIDFRKLN 62
 Db 147 Vgrllggllvpr-qspntrllpvrkpg-----tnd-----yrvpqlirevn 188
 QY 63 KATRKDHPLPFIIDHMERL-SKLTHFCFLDGYSSFSQIPVAQSDQKTTF-----TCP 115
 Db 189 krvgdlhptvnpynllsalppernwylvldkdaifclrlhptsqplfafewrdpgtgr 248
 QY 116 FGFAVRRMPFGLCNAPATFORCMMAIFSNFCENIVEV----FMDFSVYSSPDDCLSN 171
 Db 249 tggltwtclpggfknsplfdeahrdlanfrlqpgvcllgyvddlllagatkqdcleg 308
 QY 172 LDRVLQRCQDTNLVNGECKHFWMNEGIVLGHKISERGIEVDKAKVDAIDKMPYPTDIKG 231
 Db 309 tkallllesdlgyrasakkaqicrrevlylgsylrsggrwltearkkkvvgjapptakg 368
 QY 232 IRSFLGHGFGYRRFIKDF 249
 Db 369 mreftagfcrllwipgf 386

RESULT 14
 AAB73286
 ID AAB73286 standard; Protein; 1145 AA.
 XX
 AC AAB73286;
 XX
 DT 23-MAY-2001 (first entry)
 XX
 DE Retroviral protein #2 found in miniature swine.

```

XX KM Retrovirus; graft transplantation; xenotransplantation; miniature swine.
XX OS Unidentified.
XX PN US6190861-B1.
XX PD 20-FEB-2001.
XX PF 13-DEC-1996; 96US-0766528.
XX PR 14-DEC-1995; 95US-0572645.
XX PA (GEHO ) GEN HOSPITAL CORP.
XX PI Fishman JA;
XX DR WPI: 2001-256211/26.
XX N-PSDB; AAF77727.
XX PT Assessing risk of endogenous retroviruses in clinical practice and in
XX PT xenotransplantation, comprises using probe sequences derived from swine
XX PT or miniature swine retroviral genome -
XX PS Disclosure: Fig 3; 127pp; English.
XX CC The present invention relates to a method for screening a cell or tissue
XX CC for the presence or expression of a retrovirus (RV), comprising
XX CC contacting a target nucleic acid from the cell or tissue with a second
XX CC nucleic acid from the present invention (e.g. AAF77727 or a fragment
XX CC thereof). The method is useful for RV detection and to assess graft
XX CC transplantation risk. Screening of animals allows the elimination of
XX CC donors with active replication of known viruses. Inactive proviruses can
XX CC be detected and inactivated, allowing identification and elimination of
XX CC potential human pathogens derived from swine in a manner not possible in
XX CC the outbred human organ donor population and is important to the
XX CC development of human xenotransplantation.
XX SQ Sequence 1145 AA;

Query Match 14.1%; Score 189.5; DB 22; Length 1145;
Best Local Similarity 25.6%; Pred. No. 2.5e-11;
Matches 66; Conservative 44; Mismatches 119; Indels 29; Gaps 6;

OY 3 VVKLDEGIYVHNSDWSPVHSVPKKGITVVPNDKDELIPQRIITGYRMYIDFRKLN 62
DB 147 VGLLGGILLPV-GSPWNTPLIPVKP-----LND-----YRPVQDLREVN 188
OY 63 KATRKHYPLPFDHMLERL-SKLTHFCFLDGYSSFSQIPVAOSDQEKTF-----TCP 115
DB 189 KRYQDHPRTVPNYLLICLPQTSWYLVLDKDAFFCLIRHPTSPILFAFEWRDPGTGR 248
OY 116 FGFFAVRMPFGICNAPATFORCMAIFSNFCENIIVEV---FMDFSVYSSFPDCLSN 171
DB 249 TGGTLTRIPGFKNSPTLIDEALHNDLNFIRHPVLLGYVDLLIAGATKDCLEG 308
OY 172 LDVLOKCKDTNLVNGECHEFMVNEGIVLGHKISERGLVDKAKVDAIDKMPYPTDICK 231
DB 309 LKALLLESLDGYRASAKKAGLCREVTLYGSLRDGQWLTAKKTVGIPAPITAKQ 368
OY 232 IRSEFGHGFFRRFIKDF 249
DB 369 MREFLTAGTCITLWPGF 386

RESULT 15
AAR05898
ID AAR05898 standard; protein; 1784 AA.
XX AC AAR05898;
XX DT 29-NOV-1990 (first entry)

```

```

XX DE Gene product of first ORF of virus capable of inducing viraemia in
XX DE feline species encoded by clone 61E.
XX KW Viraemia; Leukaemia; FeLV-A; clone 61C; clone 61E; ds.
XX OS Feline leukaemia virus-A.
XX PN EP377842-A.
XX PD 18-JUL-1990.
XX PF 12-DEC-1989; 89EP-0122964.
XX PR 13-DEC-1988; 88US-0284139.
XX PA (HARD ) HARVARD UNIV.
XX PA (COLS ) COLARADO STATE UNIV.
XX PI Hoover EA, Mullins JT;
XX DR WPI: 1990-218326/29.
XX N-PSDB; AAO05252.
XX PT DNA encoding virus-A sub: type - producing AIDS type disease in
XX PT cats, used to test drugs and vaccines.
XX PS Claim 6; Fig 1.1-1.7; 23pp; English.
XX CC Variant of viral genome induces immunodeficiency in cats similar to
XX CC AIDS. May be used in research, especially in testing drugs and
XX CC vaccines against viraemia and and feline leukaemia viruses.
XX SQ Sequence 1784 AA;

Query Match 13.0%; Score 174.5; DB 11; Length 1784;
Best Local Similarity 23.2%; Pred. No. 2e-09;
Matches 60; Conservative 50; Mismatches 118; Indels 31; Gaps 7;

OY 5 KLDDEGIYVHNSDWSPVHSVPKKGITVVPNDKDELIPQRIITGYRMYIDFRKLNKA 64
DB 780 RMLDQGL-KPCGSPWNTPLIPVK-----PGLKD-----YRPVQDLREVNKR 821
OY 65 TRKHYPPLPFDHMLERLKLTH--FCFLDGYSSFSQIPVAOSDQEKTFCTP-----F 116
DB 822 VEDLHPRTVPNYLLICLP-SHPWYLVLDKDAFFCLIRHPTSPILFAFEWRDPGTGR 880
OY 117 GTFFAVRMPFGICNAPATFORCMAIFSNFCEN---IVEYFMDFSVYSSFPDCLSNL 172
DB 881 GGLTCTRIPGFKNSPTLIDEALHNSDLADFVRYPALVLLGYVDLLIAGATRECELG 940
OY 173 LDVLOKCKDTNLVNGECHEFMVNEGIVLGHKISERGLVDKAKVDAIDKMPYPTDICKI 232
DB 941 KALLETLGNKGYRASAKKAGLCREVTLYGSLRDGQWLTAKKTVGIPAPITAKQ 1000
OY 233 RSEFGHGFFRRFIKDF 251
DB 1001 REFLTAGTCITLWPGFAE 1019

Search completed: April 1, 2002, 22:42:41
Job time: 4388 sec

```

...


```

RESULT      1
F85074      hypothetical protein Atg4g07600 [imported] - Arabidopsis thaliana
C:Species:   Arabidopsis thaliana (mouse-ear cress)
C>Date:      16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: F85074
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor Laboratory Press
Nature 402, 769-777, 1999
A>Title:      Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488
A:Accession: F85074
A>Status:     preliminary
A:Molecule type: DNA
A:Residues:   1-630 <STO>
A:Cross-references: GB:NC_001268; NID:g7267357; PIDN:CAB81130.1; GSPDB:GN00140
C:Genetics:
A:Gene:        Atg4g07600
A:Map position: 4

Query Match          62.7%, Score 844.5; DB 2; Length 630;
Best Local Similarity 66.2%; pred. No. 1.6e-67;
Matches 159; Conservative 20; Mismatches 36; Indels 25; Gaps 2;

QY    DW-----VSVHNVPRKGGITVVVPNDKDELIPRITGTYGVYDFRLNKATRKDHYP 71
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    315 DMSKLAKPVASVQCIPKKGGITVTKNEKDELPRTITGLMCLDYYRLNAASNDRHP 374
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY    72 LPFDIMLERLSKLTFFCLDGYSSFSQIPVAOSDOEKTFTCPGTFAVRMPGLCNA 131
       ||| | | | | | : | | | | | | | | | | | | | | | | | | | | | | |
DB    375 LPTDQMELERLANHHYYCFCLDGYSGFQPIHPNDHEKTEFTFCPGTFAVERMPGLCNA 434
       ||| | | | | | : | | | | | | | | | | | | | | | | | | | | | | |

QY    132 PATFORCMATISNCNCEIVVFEMDFSYGSSPDDCSNLBRYORCDTNVLNGEKC 191
       ||| | | | | | : | | | | | | | | | | | | | | | | | | | | | | |
DB    435 PATFORCMTISDIIIEENVEVPMDFSYGSSFSCSLNLRGLTRCETNLVLNWMEGC 494
       ||| | | | | | : | | | | | | | | | | | | | | | | | | | | | | |

QY    192 HFVNVEGIYLGHKRSERGIEVDRAKVADIDMKPYPTDIGIRSFSGHGGSFYRRFIKDPFK 251
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    495 HFMVREGIILGHKHSIKSEKGLEVDKG-----CFLGHAGFYRRFIKDPFSK 536
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT      2
DBS075      probable athila transposon protein [imported] - Arabidopsis thaliana
C:Species:   Arabidopsis thaliana (mouse-ear cress)
C>Date:      16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: DBS075
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor Laboratory Press
Nature 402, 769-777, 1999
A>Title:      Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488
A:Accession: DBS075

```

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-724 <STO>
 A:Cross-references: GB:NC_001268; NID:97267363; PIDN:CAB81136.1; GSPDB:GN00140
 C:Genetics:
 A:Gene: ATAg07660
 A:Map position: 4

Query Match 62.4%; Score 841; DB 2; Length 724;
 Best Local Similarity 66.8%; Pred. No. 3.9e-67;
 Matches 155; Conservative 32; Mismatches 45; Indels 0; Gaps 0;

QY 1 KEVVKLDEGIIYVAHSDWSPVHVPKKGITVVPNDKDELIPORITTYGRMYIDPRK 60
 DB 367 KEIKLLDAGVIYIPISDSTWFPVHCVKKGITVVKNEKELLIPRITITHRACIDYRK 426
 QY 61 LNKATRKDHPLPFIIDHMLERLSKLTDFGDISSFSQIPVAOSDOEKTFTCPGTFA 120
 DB 427 LNAASRKDHPLPFTNOMLEGLANHLVNCFLDGYSGFQIPIHNDDEKTFCTCPGTFA 486
 QY 121 YRRMPFGLCNAPATFORCMAIFSNFCENIVEFMDPSVYSSFPDCLSLMDLVLRCK 180
 DB 487 YKRMPEGLCNAPATFORCMTSIFSDLEKVEYFMDPSVYSGPSSCLLNLGRVLTKE 546
 QY 181 DTNLVLNGEKCHFNVEGIVLGHKISERGIENVDAKAVDAIDKMPYPTDIKGI 232
 DB 547 ETNLVLNMEKCFYVMEKGIYVGHKISEKIEVDKIKVMMQIDPPTVMDI 598

RESULT 3
 D84513
 Probable retroelement pol polyprotein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: D84513
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 407, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: D84513
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-841 <STO>
 A:Cross-references: GB:AE002093; NID:94388825; PIDN:AAD19780.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g14040
 A:Map position: 2

Query Match 61.3%; Score 826; DB 2; Length 841;
 Best Local Similarity 62.2%; Pred. No. 1e-65;
 Matches 156; Conservative 30; Mismatches 37; Indels 28; Gaps 2;

QY 1 KEVVKLDEGIIYVAHSDWSPVHVPKKGITVVPNDKDELIPORITTYGRMYIDPRK 60
 DB 392 KEIKLLDAGVIYIPISDSTWSPVHVPKKGITVVKNEKELLIPRITITGHMCIDYRK 451
 QY 61 LNKATRKDHPLPFIIDHMLERLSKLTDFGDISSFSQIPVAOSDOEKTFTCPGTFA 120
 DB 452 LNSATRKDHPLPFIIDHMLERLSKLTDFGDISSFSQIPVAOSDOEKTFTCPGTFA 511
 QY 121 YRRMPFGLCNAPATFORCMAIFSNFCENIVEFMDPSVYSSFPDCLSLMDLVLRCK 180
 DB 512 YRRMPFGLCNAPATFORCMAIFSNFCENIVEFMDPSVYSSFPDCLSLMDLVLRCK 556
 QY 181 DTNLVLNGEKCHFNVEGIVLGHKISERGIENVDAKAVDAIDKMPYPTDIKGI 240
 DB 557 DKHLVLNMEKCFYVMEKGIYVGHKISEKIEVDKIKVMMQIDPPTVMDI 603
 QY 241 FYRRFIKIDFTK 251

DB 604 FYRRFIKIDFTK 614

RESULT 4
 T12085

reverse transcriptase homolog - fava bean (fragment)
 C:Species: Vicia faba (fava bean)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Nov-2000
 C:Accession: T12085
 R:Kinoshta, T.; Wada, H.; Masaaki, I.; Shimazaki, K.
 Submitted to the EMBL data library, September 1997
 A:Description: Retrotransposon-like cDNAs from guard cell protoplasts in Vicia faba.
 A:Reference number: 217406
 A:Accession: T12085
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-407 <KIN>
 A:Cross-references: EMBL:AB007466; NID:d1170509; PID:d1023656
 A:Experimental source: guard cell protoplasts
 C:Superfamily: pol polyprotein

Query Match 57.5%; Score 774; DB 2; Length 407;
 Best Local Similarity 70.9%; Pred. No. 1.9e-61;
 Matches 139; Conservative 29; Mismatches 28; Indels 0; Gaps 0;

QY 56 IDFRKLNKATRKDHPLPFIIDHMLERLSKLTDFGDISSFSQIPVAOSDOEKTFTCP 115
 DB 3 IDYRRLNMLATRKDHPLPFIIDHMLERLADHXYCFDGYSGYQNLAVPEDEKTTFTCP 62
 QY 116 FGTFYRRMPFGLCNAPATFORCMAIFSNFCENIVEFMDPSVYSSFPDCLSLMDLV 175
 DB 63 FGTFYRRMPFGLCNAPATFORCMTSIFSDLEKVEYFMDPSVYSGPSSCLLNLGRVLT 122
 QY 176 LORCKDTNLVLNGEKCHFNVEGIVLGHKISERGIENVDAKAVDAIDKMPYPTDIKGI 235
 DB 123 LERQESNLNLNMEKCFYVMEKGIYVGHKISEKIEVDKIKVMMQIDPPTVMDI 182
 QY 236 LGHGFYRRFIKIDFTK 251
 DB 183 LGHGFYRRFIKIDFTK 198

RESULT 5
 B96492

probable polyprotein, 77260-80472 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: B96492
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 Ansen, N.F.; Hughes, B.; Huizer, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzia,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: AB6141; MUID:21016719
 A:Accession: B96492
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-884 <STO>
 A:Cross-references: GB:AE005173; NID:g10092524; PIDN:AG12922.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: T4121.10
 A:Map position: 1

Query Match 47.8%; Score 644; DB 2; Length 884;
 Best Local Similarity 58.1%; Pred. No. 1.9e-49;

Db 475 SNFKIQMDKSEFLKLETAVLGHIIISRDGIKPNPKISAIOKYLLPPTKEIKQIFGLIGY 534
 QY 242 YRRFIKDFPK 251
 Db 535 YRKFIIPDFAR 544

RESULT 9
 534639
 pol protein - fruit fly (Drosophila ananassae) transposon Tom (fragment)
 C:Species: Drosophila ananassae
 C>Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 26-Aug-1999
 C:Accession: S34639
 R:Panda, S.; Corces, V.G.
 Submitted to the EMBL Data Library, July 1993
 A:Description: Structural and functional analysis of a Drosophila retrotransposon that
 A:Reference number: S34638
 A:Accession: S34639
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1040 <TAN>
 A:Cross-references: EMBL:224451; NID:9394703; PIDN:CAA0824.1; PID:9394705
 C:Genetics:
 A:Gene: FLYBase:Dana/Tom
 A:Cross-references: FLYBase:Fbgn0004357
 A:Mobile element: retrotransposon Tom
 C:Superfamily: pol polyprotein

Query Match 32.4%; Score 436; DB 2; Length 1040;
 Best Local Similarity 37.9%; Pred. No. 8.6e-31;
 Matches 94; Conservative 49; Mismatches 91; Indels 14; Gaps 3;
 QY 2 EVYKLLDEGIIVYHVAHSDVSPVHSVPKGGITVVPNDKDELIPORITGTYRMVIDPERKL 61
 Db 186 QVQEMLEGGIT-RESNSPYNSPTWVVPK-----PDASG-----AKRYVIDYRKL 231
 QY 62 NKATRKDHPLPFIIDHMLERLSKLTGHCFLDGSFSQIPVAVOSDQETTCPTGPTAY 121
 Db 232 NETTIDRPFIPMDIILGKLCQYFTTIDLAGRFHQIEMSESIQKTAFTKRGHYEX 291
 QY 122 RRMPEGLCNAPATFORCMAAIFSNFCENTIVEFMDPFSYGSFDDCLSNLDVLRCKD 181
 Db 292 VRRPGLRNAPATFORCMAAIFSNFCENTIVEFMDPFSYGSFDDCLSNLDVLRCKD 351
 QY 182 TNLVNGEKCHPMVNEGIVLGHKISERGIENVDAKAYDAIDKMPYPTDINGIRSLGHGCF 241
 Db 352 SNLKLQDKCEFLKKEATFLGHIVPDGIKPNPLKVEALIASYPIPTKKEIRAFILGTMGY 411
 QY 242 YRRFIKDF 249
 Db 412 YRKFIIPSY 419

RESULT 10
 534639
 retrovirus-related pol polyprotein - fruit fly (Drosophila melanogaster) retrotransposon
 N:Alternate names: reverse transcriptase
 C:Species: Drosophila melanogaster
 C>Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 16-Jun-2000
 C:Accession: A03971
 R:Salgado, K.; Kogut, W.; Matsuo, Y.; Inouye, S.; Yoshio, K.; Yuki, S.
 Nature 312, 659-661, 1984
 A:Title: Identification of the coding sequence for a reverse transcriptase-like enzyme
 A:Reference number: A93349; MUID:85061628
 A:Accession: A03971
 A:Molecule type: DNA
 A:Residues: 1-1058 <SAI>
 A:Cross-references: GB:X01472; GB:J01060; GB:J01061; NID:98142; PIDN:CAA25702.1; PID:913
 C:Genetics:
 A:Gene: FLYBase:17.6
 A:Cross-references: FLYBase:Fbgn0000004
 C:Superfamily: pol polyprotein

C:Keywords: polyprotein; reverse transcriptase

Query Match 31.7%; Score 427; DB 1; Length 1058;
 Best Local Similarity 37.9%; Pred. No. 5.6e-30;
 Matches 94; Conservative 47; Mismatches 93; Indels 14; Gaps 3;
 QY 2 EVYKLLDEGIIVYHVAHSDVSPVHSVPKGGITVVPNDKDELIPORITGTYRMVIDPERKL 61
 Db 226 QIQDMLNOGII-RTNSPNSPDIWVVPK-----QDASGKAK-----PRIVIDYRKL 271
 QY 62 NKATRKDHPLPFIIDHMLERLSKLTGHCFLDGSFSQIPVAVOSDQETTCPTGPTAY 121
 Db 272 NETTIDRPFIPMDIILGKLCQYFTTIDLAGRFHQIEMSESIQKTAFTKRGHYEX 331
 QY 122 RRMPEGLCNAPATFORCMAAIFSNFCENTIVEFMDPFSYGSFDDCLSNLDVLRCKD 181
 Db 332 LRMPEGLKNAPATFORCMAAIFSNFCENTIVEFMDPFSYGSFDDCLSNLDVLRCKD 391
 QY 182 TNLVNGEKCHPMVNEGIVLGHKISERGIENVDAKAYDAIDKMPYPTDINGIRSLGHGCF 241
 Db 392 ANLKLQDKCEFLKKEATFLGHIVPDGIKPNPKIEAIOKYLLPPTKEIKAFGLIGY 451
 QY 242 YRRFIKDF 249
 Db 452 YRKFIIPNF 459

RESULT 11
 T01842
 hypothetical protein F9D12.11 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 24-Mar-1999
 C:Accession: T01842
 R:Murray, J.; Langston, Y.; Ahrens, C.
 Submitted to the EMBL Data Library, July 1998
 A:Description: The sequence of Arabidopsis thaliana F9D12.
 A:Reference number: Z14444
 A:Accession: T01842
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1522 <MUR>
 A:Cross-references: EMBL:AF077407; NID:93319339; PID:93319351
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Map position: 4
 A:Introns: 180/2; 949/1; 1048/3
 A:Note: F9D12.11

Query Match 31.7%; Score 427; DB 2; Length 1322;
 Best Local Similarity 39.2%; Pred. No. 7.3e-30;
 Matches 91; Conservative 37; Mismatches 86; Indels 18; Gaps 1;
 QY 18 SDWSPVHSVPKGGITVVPNDKDELIPORITGTYRMVIDPERKLKATRKDHPLPFIIDH 77
 Db 557 SPWRAPMLPMLKAKKDD-----SFLCIDYGLNGQVYTKNYPPLPRIDE 598
 QY 78 MIERLSKLTGHCFLDGSFSQIPVAVOSDQETTCPTGPTAYRMRPGLCNAPATFOR 137
 Db 599 LLDQLRGATCFSKIDITSYHOIPLAEDVKTARTKRGHEFVYVMPGLNAPAAFMAR 658
 QY 138 CMAAIFSNFCENTIVEFMDPFSYGSFDDCLSNLDVLRCKDNLVNGEKCHPMVNE 197
 Db 659 LMNSVFOEFLDEFVILFIDDLIVYSKSPDEHVVHARRKYEKRLREKTLAKLSKCFWORE 718
 QY 198 GIVIGHKISERGIENVDAKAYDAIDKMPYPTDINGIRSLGHGCFYRRFIKDF 249
 Db 719 MGLTSHIVASGVSDPEKIEAIRDMRPPTNAVEIRSLFGLAGYRRFVKGF 770

RESULT 12
 A84460

Query Match	31.7%;	Score 426.5;	DB 2;	Length 1411;
Best Local Similarity	38.6%;	Pred. No. 8.8e-30;		
Matches 96;	Conservative 39;	Mismatches 95;	Indels 19;	Gaps 2;

241	FYRRFIKDF	249
	: :	
756	YYRRFVKGF	764

Query Match	31.68;	Score 425;	DB 2;	Length 1217;
Best Local Similarity	37.58;	Pred. No. 1e-29;		
Matches	94;	Conservative	46;	Mismatches 97;
				Indels 14;
				Gaps 3

[illegible]

RESULT 14
G84493
[protein not imported] - Arabidopsis thaliana

Query Match	31.1%;	Score 419.5;	DB 2;	Length 1611;
Best Local Similarity	36.9%;	Pred. No. 4.3e-29;		
Matches 92; Conservative	42;	Mismatches 96;	Indels 19;	Gaps 2

RESULT 15
B24872
not polynorbornene homopolymer - fruit fly (*Drosophila* spp.)
not polynorbornene homopolymer - fruit fly (*Drosophila* spp.)

R;Inouye, S.; Yuki, S.; Saigo, K.
Eur. J. Biochem. 154, 417-425, 1986

A:Title: Complete nucleotide sequence and genome organization of a *Drosophila* transposase
A:Reference number: A91159; MUID:86108354
A:Accession: B24872
A:Molecule type: DNA
A:Residues: 1-1059 <NO>
A:Cross-references: GB:X03431; NID:g8146; PIDN:CAB57796.1; PID:g601506
A:Note: the authors translated the codon AAT for residue 1016 as Val and GTA for residue
C:Genetics:
A:Gene: FLYBase:297
A:Cross-references: FLYBase:FBgn0000005
C:Superfamily: pol. polyprotein
C:Keywords: polyprotein

Query Match	31.0%	Score 418	DB 2	Length 1059
Best Local Similarity	37.9%	Pred. No. 3.6e-29		
Matches 94	Conservative 44	Mismatches 96	Indels 14	Gaps 3

[illegible]

Search completed: April 1, 2002, 22:44:03
Job time: 3240 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2002, 22:42:43 ; Search time 67.95 Seconds
(without alignments)
135.436 Million cell updates/sec

Title: US-09-586-106-63

Perfect score: 1347
Sequence: 1 KEVVKLDEGIIYVAHSDW.....INSLFGHGGRFRIRIDFKK 251

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	427	31.7	1058	POL3_DROME	P04323 drosophila
2	418	31.0	1059	POL2_DROME	P20823 drosophila
3	408.5	30.3	1237	POL4_DROME	P10394 drosophila
4	370.5	27.5	1035	POLX_DROME	P10401 drosophila
5	339	25.2	1333	RDPO_SCHPO	O05654 schizosacch
6	292	21.7	1745	YL51_CAEEL	P34431 caenorhabdi
7	264.5	19.6	1268	YRD6_CAEEL	O09575 caenorhabdi
8	245.5	18.2	679	POL_CAVVC	O02964 cauliflower
9	245.5	18.2	679	POL_CAVVC	O02964 cauliflower
10	245.5	18.2	679	POL_CAVVC	P33554 cauliflower
11	242	18.0	142	RRPO_OENBE	P31843 oenothera b
12	241.5	17.9	680	POL_CAVVC	O00862 cauliflower
13	238.5	17.7	674	POL_CAVVC	P03556 cauliflower
14	232.5	17.3	666	POL_FMYD	P09523 figwort mos
15	227.5	16.9	1675	POL_RTBVP	P27502 rice tungro
16	223	16.6	1886	POL_COVAV	P19199 commelina y
17	208	15.4	692	POL_SCKAV	P15629 soybean chl
18	192	14.3	659	POL_CERY	P05400 carnation e
19	191.5	14.2	1165	POL_GALV	P21414 gibdon ape
20	190.5	14.1	1046	POL_FENV1	P31792 feline endo
21	186.5	13.8	1189	POL_BAEVM	P10272 baboon endo
22	168.5	12.5	1161	POL_SEV1	P23074 simian foam
23	164.5	12.2	1204	POL_MLVFP	P26808 friend muri
24	164.5	12.2	1204	POL_MLVFP	P27401 friend muri
25	163.5	12.1	1157	POL_SEV3L	P26810 friend muri
26	159.5	11.8	1204	POL_MLVF5	P14350 human spuna
27	159	11.8	886	POL_FOAVV	P03355 moloney mur
28	151.5	11.2	1199	POL_MLVAV	P03356 akv murine
29	151.5	11.2	1196	POL_MLVAV	P11227 radiation m
30	141.5	10.5	1196	POL_MLVAV	P19560 bovine immu
31	134	9.9	1056	POL_BIV06	P19561 bovine immu
32	134	9.9	1056	POL_BIV27	P19561 bovine immu
33	133.5	9.9	1046	POL_SIVAG	P27980 simian immu

34	129.5	9.6	863	1	POL_IPHA	P04026 hamster int
35	127.5	9.5	852	1	POL_BIVAU	P25059 bovine leuk
36	125.5	9.3	852	1	POL_BIV2	P03361 bovine leuk
37	124.5	9.2	1035	1	POL_HV2NZ	P05962 human immu
38	124.5	9.2	1036	1	POL_HV2NO	P04584 human immu
39	123.5	9.2	1019	1	POL_HV2RO	P12502 simian immu
40	123.5	9.2	1035	1	POL_SIVS4	074120 human immu
41	123.5	9.2	1035	1	POL_HV2SR	P12451 human immu
42	120.5	8.9	1055	1	POL_HV2ST	P20876 human immu
43	120.5	8.9	1055	1	POL_HV2ST	P19505 simian immu
44	119.5	8.9	1022	1	POL_SIVSP	P18096 human immu
45	118.5	8.8	1009	1	POL_HV2BE	P22382 simian immu

ALIGNMENTS

RESULT 1

POL3_DROME STANDARD; PRT; 1058 AA.

AC P04323; 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RETROVIRUS-RELATED POL. POLYPROTEIN FROM TRANSPOSON 17.6 [CONTAINS:
DE PROTEASE (EC 3.4.23.-); REVERSE TRANSCRIPTASE (EC 2.7.7.49);
DE ENDONUCLEASE].
GN POL.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN
RP SEQUENCE FROM N.A.
RA MEDLINE=85061628; PubMed=6209583;
RX Salgo K., Kugimiyu W., Matsuo Y., Inouye S., Yoshiooka K., Yuki S.;
RT "Identification of the coding sequence for a reverse
transcriptase-like enzyme in a transposable genetic element in
Drosophila melanogaster";
RT Nature 312:659-661(1984).
RL
CC -I- MISCELLANEOUS: THE OPEN READING FRAME IS LOCATED IN A COPIN-LIKE
-I- TRANSPOSABLE ELEMENT CALLED 17.6.
-I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U22.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: X01472; CA25702.1; -
PIR: A03971; GNEF17.
DR HSSP: P04585; IRT1.
DR MEMOPS: A02.052; -
DR FLYBASE: FBgn0000004; 17.6.
DR InterPro: IPR001995; Asp_prot.retrov.
DR InterPro: IPR001969; Asp_protase.
DR InterPro: IPR00477; RVISE.
DR InterPro: IPR001584; Rve.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00077; rvp; 1.
DR Pfam: PF00078; rvt; 1.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;
KW Endonuclease; Transferrase; Polypeptide; Transposable element.
FT ACT_SITE 30
FT ACT_SITE 30
SQ SEQUENCE 1058 AA; 122697 MW; C893F5C4A7EIF091 CRC64;

Query Match 31.7%; Score 427; DB 1; Length 1058;

Best Local Similarity 37.9%, Pred. No. 5.9e-32;
Matches 94; Conservative 47; Mismatches 93; Indels 14; Gaps 3;

QY 2 EVKLDEGIIVHAHSMVSPVHSPKGGITVVPNDKDELIPORITGYMVDIFKL 61
D 226 QIOMLNGIIL-RSNSYNSPIWVPRK-----QDSGKOK---FIVVIDYRKL 271
QY 62 NKATRKDHVPLPFDIDHMERLSKLTGFCFLDGYSSFSQIPVAQSDQEKTFCTPGTFAY 121
D 272 NEIVGRHPIPNMDELIGKLGRCYFTTIDLAKGFHQIEMDEESVSTAFSTKSHYEX 331
QY 122 RRMFGCLNAPATFORCMMALFNSNCENIVEFMDDESFGSSPDDCLSNDRVLQRCRD 181
D 332 LRMEFGKLNAPATFORCMMNLRLPLNKHCLVYDDIIFSTSLDEHLQSLGLVEFKLAK 391
QY 182 TNVLNGECHKFHWNEGIVLGHKISERGIEVDKAKVDAIDKMPYPTDIKGRSFLGHGCF 241
D 392 ANLKLQDKCEFLKQETFTLGHVLTLPQGIKRNPEKIEAIQYPIPTKKEIKAFGLTGY 451
QY 242 YRRFIKDF 249
D 452 YRKEIPNF 459

RESULT 2

POL2_DROME

POL2_DROME

STANDARD;

PRT; 1059 AA.

AC P20825;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE RETROVIRUS-RELATED POL. POLYPROTEIN FROM TRANSPOSON 297 [CONTAINS:

DE PROTEASE (EC 3.4.23.-); REVERSE TRANSCRIPTASE (EC 2.7.7.49);

DE ENDONUCLEASE].

GN Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_Taxid=7227;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=86108354; PubMed=2417839;

RA Inoue S., Yuki S., Saigo K.;

RT "Complete nucleotide sequence and genome organization of a Drosophila

transposable genetic element, 297.";

RL Eur. J. Biochem. 154:417-425(1986).

CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U22.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@sib-sib.ch)

CC EMBL: X03431; CA557796.1; ALT_SEQ.

DR PIR: B24872, B24872.

DR HSSP: P04585; IRTI.

DR FlyBase: FBgn0000005; 297.

DR InterPro: IPR001995; Asp_prot_retrov.

DR InterPro: IPR001969; Asp_protease.

DR InterPro: IPR000477; RVTse.

DR InterPro: IPR001584; Rve.

DR Pfam: PF00665; Irv; 1.

DR Pfam: PF00077; Irv; 1.

DR Pfam: PF00078; Irv; 1.

DR PROSITE: PS00141; Asp_PROTEASE. 1.

DR Hydrolyase: Aspartyl protease; RNA-directed DNA polymerase;

DR Endonuclease: Transferrase; Polyprotein; Transposable element.

FT ACN_SITE 30 30 PROTEASE (BY SIMILARITY).

SEQUENCE 1059 AA; 123310 MW; 3905CF38E914173D CRC64;

Query Match

Best Local Similarity 31.0%; Score 418; DB 1; Length 1059;

Matches 94; Conservative 44; Mismatches 96; Indels 14; Gaps 3;

QY 2 EVKLDEGIIVHAHSMVSPVHSPKGGITVVPNDKDELIPORITGYMVDIFKL 61
D 225 QIOMLNGIIL-RESNSPNSFTWVPRK-----PDASG-----ANKYFVIDYRKL 270
QY 62 NKATRKDHVPLPFDIDHMERLSKLTGFCFLDGYSSFSQIPVAQSDQEKTFCTPGTFAY 121
D 271 NEIVGRHPIPNMDELIGKLGRCYFTTIDLAKGFHQIEMDEESVSTAFSTKSHYEX 330
QY 122 RRMFGCLNAPATFORCMMALFNSNCENIVEFMDDESFGSSPDDCLSNDRVLQRCRD 181
D 331 LRMEFGKLNAPATFORCMMNLRLPLNKHCLVYDDIIFSTSLDEHLNSIOLVETKIALD 390
QY 182 TNVLNGECHKFHWNEGIVLGHKISERGIEVDKAKVDAIDKMPYPTDIKGRSFLGHGCF 241
D 391 ANLKLQDKCEFLKKEANFLGHVLTLPQGIKRNPKIKVAISVPIPTKKEIKAFGLTGY 450
QY 242 YRRFIKDF 249
D 451 YRKEIPNF 458

RESULT 3

POL4_DROME

POL4_DROME

STANDARD;

PRT; 1237 AA.

AC P10394;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE RETROVIRUS-RELATED POL. POLYPROTEIN FROM TRANSPOSON 412 [CONTAINS:

DE PROTEASE (EC 3.4.23.-); REVERSE TRANSCRIPTASE (EC 2.7.7.49);

DE ENDONUCLEASE].

GN Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_Taxid=7227;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=86274717; PubMed=2426108;

RA Yuki S., Inoue S., Ishimaru S., Saigo K.;

RT "Nucleotide sequence characterization of a Drosophila

retrotransposon, 412.";

RL Eur. J. Biochem. 158:403-410(1986).

CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U22.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@sib-sib.ch)

CC EMBL: X04132; CA227750.1; -

DR PIR: D29349; GNEF42.

DR HSSP: P03366; IHG3.

DR FlyBase: FBgn0000006; 412.

DR InterPro: IPR001995; Asp_prot_retrov.

DR InterPro: IPR001969; Asp_protease.

DR InterPro: IPR000477; RVTse.

DR InterPro: IPR001584; Rve.

DR Pfam: PF00665; Irv; 1.

DR Pfam: PF00077; Irv; 1.

DR Pfam: PF00078; Irv; 1.

DR PROSITE: PS00141; Asp_PROTEASE. 1.


```

RT fission yeast.
RL Gene 131:135-139(1993).
RL (2)
RP SEQUENCE FROM N.A. (SPAC26A3.13C).
RC STRAIN-972;
RA McLean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A. (SPAC27E2.08).
RC STRAIN-972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN (4)
RP SEQUENCE FROM N.A. (SPAC9.04).
RC STRAIN-972;
RA Weiler H., Duesterhoeft A., Lyne M.H., Rajandream M.A., Barrell B.G.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN (5)
RP SEQUENCE FROM N.A. (SPAC9B6.02C).
RC STRAIN-972;
RA Wood V., Rajandream M.A., Barrell B.G., Volckaert G.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN (6)
RP SEQUENCE OF 120-1333 FROM N.A. (SPAC167.08).
RC STRAIN-972;
RA Rieger M., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: REGIONALLY, TO PROTEASE, REVERSE TRANSCRIPTASE,
CC RNASE H AND INTEGRASE FROM OTHER RETROTRANSPOSONS AND
CC RETROVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L10324; AAA91215.1; -
DR EMBL: Z69240; CAA93236.1; -
DR EMBL: Z98978; CAB11682.1; -
DR EMBL: A1121764; CAB57422.1; -
DR EMBL: AL049769; CAB42363.1; -
DR EMBL: AL035248; CAB58169.1; -
DR MEROPS: A02.051; -
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR000477; RVTse.
DR InterPro: IPR001584; Rve.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00078; rvt; 1.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW transposable element.
FT SIMILAR 1 400 TO CAPSID PROTEIN FROM RETROVIRUSES.
FT VARIANT 205 206 L->P (IN SPAC27E2.08).
SQ SEQUENCE 1333 AA; 154932 MW; 2F45408238DC1DA CRC64;

```

Query Match 25.2%; Score 339; DB 1; Length 1333;
 Best Local Similarity 33.5%; Pred. No. 1,3e-23;
 Matches 75; Conservative 41; Mismatches 90; Indels 18; Gaps 1;

```

QY 23 PVLSVRKKGITVVPNDKDELIPQRIITGRVVIDFRKLKATRKDHYPLPFIIDHMLERL 82
DB 451 IIVIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
DB 451 -LRMVVDYKPLKKYVKNYIPNIPPLIIQLLAKI 492
QY 83 SKLTHCFIDGSSFSQIPVAOSDQKTTFTCPGTFAYRRMPGLCNATATQRCMAAI 142
DB 493 QSGTITTKLKLKSAHLIRKGDENHLARCPGVFEYLYMPGISTADAHQYFLNTI 552
QY 143 FSNRCNIEVFNMFDSVYSGSFDCLSNDRVLQORCKDNTNLVNGCKHFVMEGIVLG 202
DB 553 LGRAKESHVVCYMDLILHKSSEHKKHKKVLDLQKLKNNLIIINQAKCFHOSQVCFIG 612

```

```

QY 203 HKISERGIEVDKAKYDAIDKMPYPTDIKISRFSLGHGFRRFI 246
DB 613 YHISERGFTPCQENIDKYLQWKOPKRRKELRQFLGSVNYLRFI 656

```

RESULT 6

```

YL51_CAEEL
ID YL51_CAEEL STANDARD; PRT; 1745 AA.
AC P34431.
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL PROTEIN F44E2.1 IN CHROMOSOME III (FRAGMENT).
GN F44E2.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL NZ;
RA Anderson K.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RETROVIRAL-TYPE PROTEASE/REVERSE TRANSCRIPTASE/
CC ENDONUCLEASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L23646; AAA28035.1; -
DR HSSP: P04585; IRT1.
DR WormPep: F44E2.1; CE00542.
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR000477; RVTse.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR001878; ZnF_CCHC.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00078; rvt; 1.
DR Pfam: PF00098; zfc_CCHC; 1.
DR SMART: SM00343; ZnF_C2HC; 1.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW Hypothetical protein; Hydrolase; Aspartyl protease;
KW Transferrase; RNA-directed DNA polymerase.
FT ACT_SITE 382 382
FT NON_TER 1745 1745
FT PROTEASE (BY SIMILARITY).
SQ SEQUENCE 1745 AA; 198408 MW; 7C7457EECB19329 CRC64;

```

Query Match 21.7%; Score 292; DB 1; Length 1745;
 Best Local Similarity 29.1%; Pred. No. 4.2e-19;
 Matches 73; Conservative 40; Mismatches 100; Indels 38; Gaps 3;

```

QY 1 KEVYKLIDEGITIVHNSDWSPVHSVPKKGITVVPNDKDELIPQRIITGRVVIDFRK 60
DB 678 KTIQDKMLNOKVT-RESKSPMSSPVVLVKKKG-----SIRCIDYRK 718
QY 61 LNKATRKDHYPLPFIIDHMLERLSKLTHCFIDGSSFSQIPVAOSDQKTTFTCPGTF 120
DB 719 VNKVYKNNANHPRLPIETLQSLACKKIYTFDMAGWQIPLDEKSEITAFAGSLFE 778
QY 121 YRMPFGICNAPATFGRCMAITSNFCENIEVFNMFDSVYSGSFDCLSNDRVLQORCK 180
DB 779 WNVLPFGLIVSPALFQGTMEIIEIDLGVCAYVVDLLASKMQEHNLDVKEALRRIR 836
QY 181 DTNLVNGCKHFVNGIGIYLGKHSRGIEVDKAKYDAIDKMPYPTDIKISRFSLGHG 240
DB 839 KSGKILRASKCHLAKKVEEYLGKVTLLDGVETQ-----SFLGLVG 879

```


QY 230 KGI RSLG 237

ID	STANDARD	PR1	0/9 AA.
AC	P03554;		
DT	21-JUL-1986 (Rel. 01, created)		
DT	21-JUL-1986 (Rel. 01, last sequence update)		
DT	20-AUG-2001 (Rel. 40, last annotation update)		
DE	ENZYMATIC POLYPROTEIN [CONTAINS: ASPARTIC PROTEASE (EC 3.4.23.-)]		

GN Cauliflower mosaic virus (strain Strasbourg) (CaMV)
OS Viruses: Retroid viruses: Caulimovirus
OC V.

RN	[1]
RP	SEQUENCE FROM N.A.

RA Franck A., Guilley H., Jonard G., Richards K., Hirth L.
RT "Nucleotide sequence of cauliflower mosaic virus DNA.",
RI Cell 31:285-294(1980)

CC -1- SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS ORF V.
CC -1- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute.

CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/commercial>)

CC -----
DR EMBL; V00141; CAA23460.1; -
DR PDB; A04150; OOCVE

```
DR      MEROPS; A03.001; -,
DR      InterPro; IPR000588; Peptidase_A3.
```

```
DR Pfam; PF00078; rvt; 1.
DR Pfam; PF02160; peptidase_A3; 1.
DR Pfam; PF00731; GAT_TNCT137
```

Accession	Protein	Length	Accession	Protein	Length
KW	Transferase; Endonuclease; Polyprotein.				
FT	ACT_SITE	45			

FT	FT	FT	FT
260	620	TO RETROVIRUS GAG/POL DNA POLYMERASE DOMAIN.	
SIMILAR			

Query Match	18.2%	Score	245.5	DB 1	Length	679
Best Local Similarity	29.0%	Preced	36.15			

OY	Matches	72;	Conservative	49;	Mismatches	90;	Indels	37;	Gaps	8
1	KEVKKLLDEGIILYYHAHSDWSPVHSVPKKGITTVENDKDELIPRIITGYRMVIDFRK									60

```
Db      264 KQIKELLDKVI-----KPKSK-PHMAPAEVNNNEAKRRGKK-----RMVVNYKA 308
```

```

Db      309 MNKATVGDAYNLPNKDELTLLIRGKKIESSFDPKSGFWQVLLDQESRPLTAFTCGGHE 368
      :||| | | | | : | : | | | | : : : : | ||| | :

```

Dd 369 WNVVPEGLKQAPSIQRHNDEAFRYRKFC---CVYVDILVFSSNNEEDHLLAHVAMITLQ 424

```
QY      178 RCDNTNLVLNGEKECHFMVNEGIVLGHKISERGIEVDKAK-----VDAIDKMPP-TDI 229
```

OY 230 KGIKRSFLG 237
 Db 478 KQIQRFGL 485

RESULT 11
 ID RPO_OENBE STANDARD; PRT; 142 AA.

AC P31843;
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RNA-DIRECTED DNA POLYMERASE HOMOLOG (REVERSE TRANSCRIPTASE HOMOLOG).
 OS Oenothera lutea (Bertero's evening primrose).
 CC Mitochondrion.
 CC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 CC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
 CC Euphorbia: Myrtales: Onagraceae: Oenothera.
 CC Euphorbia: Myrtales: Onagraceae: Oenothera.
 CC NCBI_TaxID=3950;

RP SEQUENCE FROM N.A.
 RA Schuster W., Brennick A.;
 RT "Plastid, nuclear and reverse transcriptase sequences in the
 RT mitochondrial genome of Oenothera: is genetic information transferred
 RT between organelles via RNA?";
 RT EMBO J. 6:2857-2863(1987).
 RL -1 SIMILARITY: TO REVERSE TRANSCRIPTASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X06034; CA29429.1; -
 DR PIR: S01101; RROBHM.
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF00078; rvt. 1.
 DR Transferrase: RNA-directed DNA polymerase: Mitochondrion.
 SQ SEQUENCE 142 AA: 16799 MW: 859306FA726FA80C CRC64;

Query Match 18.0%; Score 242; DB 1; Length 142;
 Best Local Similarity 37.9%; Pred. No. 1e-15;
 Matches 50; Conservative 32; Mismatches 46; Indels 4; Gaps 2;

OY 53 RMVIDFRKLNKATKRDHYPLPFDHMLERLSTLHFCFLDGYSSFSQIPVAGSDQKTTT 112
 Db 7 RMCIDRALTKYTKKPIPRVDLFDLQATWTFKLDLSRGWQVRIAGDEPKTTC 66
 OY 113 TCPEGTFAVRMPFGICNAPATFORCMAIFSNFCENIVEYMDP---FSVYSSPDCL 169
 Db 67 VRRYGSFEFRVMPFGICNAPATFORCMAIFSNFCENIVEYMDP---FSVYSSPDCL 126
 OY 170 SNIDRVLRQCD 181
 Db 127 KHL-RVRESKE 137

RESULT 12
 ID POL_CAMVN STANDARD; PRT; 680 AA.

AC 000962;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ENZYMOLOGICAL POLYMERASE [CONTAINS: ASPARTIC PROTEASE (EC 3.4.23.-);
 DE ENDONUCLEASE; REVERSE TRANSCRIPTASE (EC 2.7.7.49)].
 OS Cauliflower mosaic virus (strain NY8153) (CamV).

CC Viruses: Retroid viruses; Caulimovirus.
 CC NCBI_TaxID=31557;

RP SEQUENCE FROM N.A.
 RA Chenault K.D., Steffens D.L., Melcher U.K.;
 RT "Nucleotide sequence of cauliflower mosaic virus isolate NY8153."
 RT Plant Physiol. 100:542-545(1992).
 RL -1 SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS ORF V.
 CC -1 SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.
 CC -1 SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M9541; AAA6358.1; -
 DR HSP: P04585; 1KLM.
 DR MEROPS: A03.001;
 DR InterPro: IPR000588; Peptidase_A3.
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF02160; Peptidase_A3; 1.
 DR Pfam: PF00078; rvt. 1.
 DR PRINTS: PR00731; CAULIMOVIRUS.
 KW Hydrolyase, Aspartyl protease: RNA-directed DNA polymerase;
 KW Transferrase: Endonuclease; Polypeptide.
 FT ACT SITE 46 46 BY SIMILARITY.
 FT SIMILAR 41 131 TO RETROVIRUS GAG/POL PROTEASE DOMAIN.
 FT SIMILAR 261 621 TO RETROVIRUS GAG/POL DNA POLYMERASE
 FT DOMAIN.
 SQ SEQUENCE 680 AA: 78665 MW: FCE02E09647C2221 CRC64;

Query Match 17.9%; Score 241.5; DB 1; Length 680;
 Best Local Similarity 27.0%; Pred. No. 7.1e-15;
 Matches 69; Conservative 47; Mismatches 87; Indels 53; Gaps 7;

OY 1 KEVKKLDEGLTYVANSWSPVSHVSKGIGTYVNDKDELIPRLIT-----GV 52
 Db 265 KQIKELLDLKYI-----KPSKSPHMAPPAFLVNNENGRGK 301
 OY 53 RMVIDFRKLNKATKRDHYPLPFDHMLERLSTLHFCFLDGYSSFSQIPVAGSDQKTTT 112
 Db 302 RMVNVYKAMKATYGDVNLNPKDELTLIRGKKIFSSFCCKSGFWQLDQSRPLTAF 361
 OY 113 TCPEGTFAVRMPFGICNAPATFORCMAIFSNFCENIVEYMDP---FSVYSSPDCL 169
 Db 362 TCPEGTFAVRMPFGICNAPATFORCMAIFSNFCENIVEYMDP---FSVYSSPDCL 169
 OY 170 SNIDRVLRQCDTULVNGECKHVMNGBIYLGKISERGIEYVKAR-----VDAIDK 222
 Db 418 LHMVMILOKQMOHGLILSKKAQ-----LFKKKIFLGLEIDEGHKKQGHLEHINK 470
 OY 223 MYP-TDIKGRSEFLG 237
 Db 471 FPDITLEDKQKQIRFLG 486

RESULT 13
 ID POL_CAMVN STANDARD; PRT; 674 AA.

AC 003556;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ENZYMOLOGICAL POLYMERASE [CONTAINS: ASPARTIC PROTEASE (EC 3.4.23.-);
 DE ENDONUCLEASE; REVERSE TRANSCRIPTASE (EC 2.7.7.49)].
 OS Cauliflower mosaic virus (strain D/H) (CamV).
 CC Viruses; Retroid viruses; Caulimovirus.

OX NCBI_TaxID=10645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=83106468; PubMed=7152260;
 RX Balazs E., Guillely H., Jonard G., Richards K.:
 RT "Nucleotide sequence of DNA from an altered-virulence isolate D/H of
 RL the cauliflower mosaic virus.";
 CC Gene 19:239-249(1982).
 CC -1- SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS ORF V.
 CC -1- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M10376; AAA46350.1; ALT-INT.
 DR PIR: A04159; Q0CV5.
 DR HSSP: P04585; 1KLM.
 DR MEROPS: A03.001; -
 DR InterPro: IPR000588; Peptidase_A3.
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF02160; Peptidase_A3; 1.
 DR Pfam: PF00078; tvc; 1.
 DR PRINTS: PR00731; CAULIMOPTASE.
 DR Hydrolase: Aspartyl protease; RNA-directed DNA polymerase;
 DR Transferrase; Endonuclease; Polypeptide.
 DR ACT_SITE 47
 FT SIMILAR 40 130 TO RETROVIRUS GAG/POL PROTEASE DOMAIN.
 FT SIMILAR 260 620 TO RETROVIRUS GAG/POL DNA POLYMERASE
 FT SEQUENCE 674 AA; 78164 MW; E00AE3222D349E29 CRC64;
 Query Match 17.7%; Score 238.5; DB 1; Length 674;
 Best Local Similarity 28.6%; Pred. No. 1.3e-14;
 Matches 71; Conservative 49; Mismatches 91; Indels 37; Gaps 8;
 OY 1 KEVKKLLDEGIIVNVAISDWSPVHSVPKKGITVVPNDKDELIPQILITGYRWIDFRK 60
 DB 259 KQIKELLDLGLIT-PSKSQMSPAF-----LVNEAERRRKK- -RMVNYKA 303
 OY 61 LNKATRKDHPLPEIDHMLERLSKLTNFCFLDGYSSFSQIPVAOSDOEKTFTCPGTF 120
 DB 304 MNKATVGDAYNPKNDELTLTGKRIKIFSSFDCKSGFWYVLDQESRPLAFTCPGCHYE 363
 OY 121 YRRMPGSLCNAPATFORCM---MAISNFCENIYEVFMDFSYVSSFDCLSLNDRVLQ 177
 DB 364 MNVVPFGLKQAPSIQFQHMDEAFVRKFC---CYVVDILVFSNNEEDHLHVMILQ 419
 OY 178 RCKDTLVNNGECKHFMVNGIYLGKISRGIEVDKAK-----VDLIDKMPY-TDI 229
 DB 420 KQNGHILSKKKAQ-----LFFKKINFLGLEIDEGTHKPGCHLEHINKPPTLEDK 472
 OY 230 KGIKSLFG 237
 DB 473 KGIKSLFG 480
 RESULT 14
 POL_FMYD STANDARD: PRT; 666 AA.
 AC P09523;
 DT 01-MAR-1989 (Rel. 10; Created)
 DT 01-MAR-1989 (Rel. 10; Last sequence update)
 DT 20-AUG-2001 (Rel. 40; Last annotation update)
 DE ENZYMOLOGICAL PROTEIN [CONTAINS: ASPARTIC PROTEASE (EC 3.4.23.-);
 DE ENDOUCLEASE; REVERSE TRANSCRIPTASE (EC 2.7.7.49)].
 GN V.

OS Figwort mosaic virus (strain Dxs) (FMV).
 OC Viruses; Retroid viruses; Caulimovirus.
 OX NCBI_TaxID=10650;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=88040466; PubMed=3671088;
 RX Richins R.D., Scholthof H.B., Shephard R.J.:
 RT "Sequence of figwort mosaic virus DNA (caulimovirus group).";
 RL Nucleic Acids Res. 15:8451-8466(1987).
 CC -1- SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS ORF V.
 CC -1- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A3.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X06166; CAA29527.1; -
 DR PIR: S01283; S01283.
 DR HSSP: P04585; 1RT2.
 DR MEROPS: A03.001; -
 DR InterPro: IPR000588; Peptidase_A3.
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF02160; Peptidase_A3; 1.
 DR Pfam: PF00078; tvc; 1.
 DR PRINTS: PR00731; CAULIMOPTASE.
 DR Hydrolase: Aspartyl protease; RNA-directed DNA polymerase;
 DR Transferrase; Endonuclease; Polypeptide.
 DR ACT_SITE 54
 FT SEQUENCE 666 AA; 77081 MW; E65BC57D8FD2CA0F CRC64;
 Query Match 17.3%; Score 232.5; DB 1; Length 666;
 Best Local Similarity 28.6%; Pred. No. 4.8e-14;
 Matches 70; Conservative 52; Mismatches 92; Indels 31; Gaps 7;
 OY 1 KEVKKLLDEGIIVNVAISDWSPVHSVPKKGITVVPNDKDELIPQILITGYRWIDFRK 60
 DB 257 KQIKELLDLGLIT-PSKSQMSPAF-----LVNEAERRRKK- -RMVNYKA 301
 OY 61 LNKATRKDHPLPEIDHMLERLSKLTNFCFLDGYSSFSQIPVAOSDOEKTFTCPGTF 120
 DB 302 INKATIGDSHNLNPMOELLTLGKRIKIFSSFDCKSGFWYVLDQESOKLTAFTCPGCHY 361
 OY 121 YRRMPGSLCNAPATFORCMMAISNFCENIYEVFMDFSYVSSFDCLSLNDRVLQ 180
 DB 362 WKVVPFGLKQAPSIQFQHMOTAL-NGADKFCMYVDIIVFSNSELHYNHVAVALIVE 420
 OY 181 DTNLVINGECKHFMVNGIYLGKISRGIEVDKAK-----VDLIDKMP-YPTDIKI 222
 DB 421 KYGILLSKKKAQ-----LFFKKINFLGLEIDEGTHKPGCHLEHINKPPTLEDK 473
 OY 233 RSLFG 237
 DB 474 RSLFG 478
 RESULT 15
 POL_RTBVP STANDARD: PRT; 1675 AA.
 AC P27502; P27528;
 DT 01-AUG-1992 (Rel. 23; Created)
 DT 01-AUG-1992 (Rel. 23; Last sequence update)
 DT 20-AUG-2001 (Rel. 40; Last annotation update)
 DE POLYPROTEIN (P194 PROTEIN) [CONTAINS: COAT PROTEIN; PROTEASE
 DE (EC 3.4.23.-); REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H
 DE (EC 3.1.26.4)].
 OS Rice tungro bacilliform virus (isolate Philippines) (RTBV).
 OC Viruses; Retroid viruses.

OX NCBI_TaxID=10655;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91252204; PubMed=2041739;
 RA Hay J.M., Jones M.C., Blakebrough M.L., Dasgupta I., Davies J.W.,
 RA Hull R.;
 RT "An analysis of the sequence of an infectious clone of rice tungro
 RT bacilliform virus, a plant pararetrovirus."
 RL Nucleic Acids Res. 19:2615-2621(1991).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 503-526.
 RX MEDLINE=92024093; PubMed=1926781;
 RA Ou R.D., Bhattacharya M., Laco G.S., de Kochko A., Rao B.L.S.,
 RA Kaniewska M.B., Elmer J.S., Rochester D.E., Smith C.E.,
 RA Beachy R.N.;
 RT "Characterization of the genome of rice tungro bacilliform virus:
 RT comparison with Comelina yellow mottle virus and caulimoviruses."
 RL Virology 185:354-364(1991)
 CC -1- SIMILARITY: WITH COMELINA YELLOW MOTTLE VIRUS POLYPROTEIN, AND
 CC WITH CAULIMOVIRUSES ORF V.
 CC -1- SIMILARITY: WITH RETROVIRAL POL/GAG POLYPROTEINS.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U33.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL: X57924; CAA00997.1; -;
 DR EMBL: M65026; AAB03094.1; -;
 DR PIR: C40785; C40785.
 DR MEROPS: A03.002; -;
 DR InterPro: IPR001995; Asp_prot_retrov.
 DR InterPro: IPR001699; Asp_protease.
 DR InterPro: IPR000477; RVase.
 DR InterPro: IPR01878; ZnF_CCHC.
 DR Pfam: PF000077; rvp; 1.
 DR Pfam: PF00078; rvt; 1.
 DR Pfam: PF00098; zf-CCHC; 1.
 DR SMART: SM00343; ZnF_C2HC; 1.
 DR PROSITE: PS00141; Asp_PROTEASE; 1.
 DR PolyProtein: RNA-directed DNA polymerase; Hydrolase;
 KW Aspartyl protease; Transferase; Nucleotidyltransferase; Coat protein;
 KW Endonuclease.
 KW
 FT CHAIN 503 ? 33 KDA COAT PROTEIN.
 FT ZN_FING 772 789 POTENTIAL.
 FT ACT_SITE 987 987 PROTEASE (BY SIMILARITY).
 FT CONFLICT 1292 1292 D -> E (IN REF. 2).
 FT CONFLICT 1630 1630 Y -> H (IN REF. 2).
 FT CONFLICT 1635 1635 S -> P (IN REF. 2).
 SQ SEQUENCE 1675 AA; B24D7181463E466F CRC64;

 Query Match 16.9%; Score 227.5; DR 1; Length 1675;
 Best Local Similarity 26.0%; Pred. No. 4.2e-13;
 Matches 68; Conservative 49; Mismatches 110; Indels 35; Gaps 6;

 OY 1 KEVVKLLDEGIIVVAHSDWSPVHSVKKGGIT-----VVPNDKDLIPQRIITGY 52
 DB 1201 KQIKELLDNKLT-----KKADPTCRHRTAFTVRNHSEVAQK----- 1239

 OY 53 RNVIDFKLKKATRKQHYPLPFIIDHMLERLSKLTFCFLDYSSESQIPVAQSDQEKTF 112
 DB 1240 RIVYNNKRLNDNMHTDPFNPHKISMINTLOKANIFSKFDLKAGFHHMKLKDKFDWTF 1299

 OY 113 TCPEGTFAYRRMPGCLNAPATFORCMATFENSCENIVFEMDPSVYSSPDDCLSNL 172
 DB 1300 TCSEGLYTNVVCPRGIANACAFORFQESFGDL--KFAITYIDDIILASNNKEKIEHL 1357

 OY 173 DRYLORCKDTNLVNGEKHFVNEGIVLGHKISERGIEVDKAKVDAI---DKMPYPTDI 229

DB 1358 KIFENRVEVGCVLSSKSKMFLKEVEYLGVEIKKGKISLOPHYDKIKKFKNKLT-L 1416
 OY 230 KGIRSFLLGHGFGYRRPDKFTK 251
 DB 1417 KCLQAVLGLLNVARGVYIKDLK 1438

Search completed: April 1, 2002, 22:47:31
 Job time: 288 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2002, 22:13:28 ; Search time 114.44 Seconds
(without alignments)
320.817 Million cell updates/sec

Title: US-09-586-106-63
Perfect score: 1347
Sequence: 1 KEVVKLDEGIIYVHNSDW.....IRSEFGNGCFYRREIKDFTK 251

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 segs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1001	74.3	1864	10	Q9FZN9
2	981	72.8	1862	10	Q9SHN5
3	976	72.5	1586	10	Q9LPI1
4	947	70.3	897	10	Q9LHC0
5	946	70.2	1799	10	Q9SHM3
6	841	62.4	724	10	Q9M0R8
7	826	61.3	841	10	Q9ZPT5
8	774	57.5	407	10	Q92103
9	644	47.8	884	10	Q9C803
10	549	40.8	622	10	Q9SKC0
11	472	35.0	565	10	Q9LJ78
12	455	33.8	1062	5	Q9VZ16
13	453.5	33.7	1240	10	Q9T0B7
14	447	33.2	1504	10	Q9LKH4
15	437.5	32.5	1504	10	Q9LKH4
16	434	32.2	1150	5	Q96740
17	427	31.7	1322	10	Q9LJ90
18	426.5	31.7	1411	10	Q9SJ92
19	425	31.6	1217	5	Q46115

20	422	31.3	681	5	Q9VMU9	Q9VMU9 drosophila
21	422	31.3	1037	5	Q9NHF7	Q9NHF7 drosophila
22	420.5	31.2	1587	10	Q9SLC0	Q9SLC0 oryza sativ
23	419.5	31.1	1611	10	Q9SK57	Q9SK57 arabidopsis
24	417.5	31.0	901	10	Q9ZPE8	Q9ZPE8 arabidopsis
25	414.5	30.8	1484	10	Q9XE85	Q9XE85 sorghum bic
26	414	30.7	1188	5	Q76326	Q76326 drosophila
27	409	30.4	1060	5	Q17318	Q17318 ceratilis c
28	407.5	30.3	949	10	Q9XE44	Q9XE44 arabidopsis
29	407.5	30.3	1499	10	Q9LHM4	Q9LHM4 arabidopsis
30	404.5	30.0	924	10	Q9XEB8	Q9XEB8 sorghum bic
31	404	30.0	1328	10	Q9ZVR4	Q9ZVR4 arabidopsis
32	403	29.9	1494	5	Q94885	Q94885 drosophila
33	401.5	29.8	1265	3	Q9VUD8	Q9VUD8 alternaria
34	400.5	29.7	976	10	Q9SIX7	Q9SIX7 arabidopsis
35	399.5	29.7	871	10	Q64892	Q64892 ananas com
36	398	29.5	1901	10	Q9AUJ6	Q9AUJ6 oryza sativ
37	397.5	29.5	1021	10	Q9LJ23	Q9LJ23 oryza sativ
38	396.5	29.4	1524	10	Q9LWJ0	Q9LWJ0 oryza sativ
39	394.5	29.3	2017	10	Q9AYB6	Q9AYB6 oryza sativ
40	394	29.3	2030	10	Q9FED6	Q9FED6 oryza sativ
41	393.5	29.2	1221	3	Q05350	Q05350 saccharomyc
42	393.5	29.2	1547	3	Q99315	Q99315 saccharomyc
43	393	29.2	1783	10	Q9AWS5	Q9AWS5 oryza sativ
44	393	29.2	1867	10	Q9AUD4	Q9AUD4 oryza sativ
45	392.5	29.1	894	10	Q9FMC7	Q9FMC7 oryza sativ

ALIGNMENTS

RESULT 1
ID Q9FZN9 PRELIMINARY; PRT; 1864 AA.

AC Q9FZN9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE RETROELEMENT POL POLYPROTEIN-LIKE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA.
RA Kaneke T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
TRANSCRIPTASE)
CC EMBL; AB028613; BAB10790.1; -
DR EMBL; AB024037; BAB10790.1; JOINED.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00078; rvt; 1.
KW polyprotein; RNA-directed DNA polymerase.
SO SEQUENCE 1864 AA; 212130 MW; 31271031F77DBCF CRC64;

Query Match 74.3%; Score 1001; DB 10; Length 1864;
Best Local Similarity 73.3%; Pred. No. 6.3e-86;
Matches 184; Conservative 28; Mismatches 39; Indels 0; Gaps 0;

QY 1 KEVVKLDEGIIYVHNSDWSPVHSVKKGGIYVVPDKKELLIPQRTITTYRNVAVIFRK 60
DB 976 KETLKLDAAGVYIPISDSTWSPVHYVKKGMVTVKNSKDELLPRTTTHRCIDIRK 1035
QY 61 LKAKRKHHYPLPFDHMLERSLTHFCPLDGYSSFOIVPAOSDOKRTFTCPFGTFA 120
||:||||:||||| ||||| :||||| ||| :||||| |||||

```

Db 1036 LNAASRKHFPLPFDIDMLERLANHPYCCLDGYSGFQIPHPNDQKTEFTCPGTFA 1095
QY 121 YRRMPFGLCNAPATFORCMAIFSNFCENIVEFMDDFSVYSSFPDCLSNDRYLQCK 180
Db 1096 YRRMPFGLCNAPATFORCMTSIFSDLEIEMVEFMDDFSVYSSFPDCLSNDRYLQCK 1155
QY 181 DTNVLNCKEKFHVNNEGIVGKHISERGIEVDKAKYDAIDKMPYPTDIKIGIRSLGHG 240
Db 1156 ETNVLNCKEKFHVNNEGIVGKHISERGIEVDKAKYDVMMQLODPKTVKDIRSLGHAG 1215
QY 241 FYRRFIKDFPK 251
Db 1216 FYRRFIKDFPK 1226

RESULT 2
Q9SHM5 PRELIMINARY; PRT; 1862 AA.
AC Q9SHM5;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE F7F22.15.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Chao Q., Shinn P., Brooks S., Buehler E., Dunn P., Khan S., Kim C.,
RA Walker M., Brooks S., Altafi H., Araujo R., Conn L., Conway A.B.,
RA Gonzalez A., Hansen N.F., Huizar U., Kremenetskaia I., Lenz C., Li J.,
RA Liu S., Lutos S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V.,
RA Yu G., Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
RA Genomic sequence for Arabidopsis thaliana BAC F7F22 from chromosome
RT I."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
CC TRANSCRIPTASE).
DR EMBL: AC007534; AAF24529.1; -.
DR InterPro: IPR001584; RVE.
DR InterPro: IPR004477; RVTSE.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00078; rve; 1.
DR RNA-directed DNA polymerase.
KW SEQUENCE 1862 AA; 212491 MW; 768944AA61EA519B CRC64;
SO

Query Match 72.8%; Score 981; DB 10; Length 1862;
Best Local Similarity 71.3%; Pred. No. 5e-84; Mismatches 41; Indels 0; Gaps 0;
Matches 179; Conservative 31;

QY 1 KEVVLLDDEGIIYVHAHSDWSPVSHVPRKGGITVVPNDKDELIPORITTYGRMYIDFK 60
Db 975 KEILLLDAGVITYPISDSVWSPVHCVPKDKGMIVYKNEKDELIPRTITGHRMCIDYRK 1034
QY 61 LNKATRKDHYPPLPFDHMLERLSKLTTHFCFLDGYSSFSQIPVAOSDOEKTTPCFFGFA 120
Db 1035 LNAASRKHFPLPFDIDMLERLANHPYCCLDGYSGFQIPHPNDQKTEFTCPGTFA 1094
QY 121 YRRMPFGLCNAPATFORCMAIFSNFCENIVEFMDDFSVYSSFPDCLSNDRYLQCK 180
Db 1096 YRRMPFGLCNAPATFORCMTSIFSDLEIEMVEFMDDFSVYSSFPDCLSNDRYLQCK 1155
QY 181 DTNVLNCKEKFHVNNEGIVGKHISERGIEVDKAKYDAIDKMPYPTDIKIGIRSLGHG 240
Db 1156 ETNVLNCKEKFHVNNEGIVGKHISERGIEVDKAKYDVMMQLODPKTVKDIRSLGHAG 1215
QY 241 FYRRFIKDFPK 251
Db 1216 FYRRFIKDFPK 1225

```

```

RESULT 3
Q9LPB1 PRELIMINARY; PRT; 1586 AA.
AC Q9LPB1;
DT 01-OCT-2000 (TRENBLREL. 15, Created)
DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)
DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE T32E20.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altafi H., Bel O., Chin C., Chou J., Choi E., Conn L.,
RA Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RA Genomic sequence for Arabidopsis thaliana BAC T32E20 from chromosome
RT I."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
CC TRANSCRIPTASE).
DR EMBL: AC020646; AAF79809.1; -.
DR InterPro: IPR001584; RVE.
DR InterPro: IPR004477; RVTSE.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00078; rve; 1.
DR RNA-directed DNA polymerase.
KW SEQUENCE 1586 AA; 180875 MW; A9D3917EAFDAD790 CRC64;
SO

Query Match 72.5%; Score 976; DB 10; Length 1586;
Best Local Similarity 71.3%; Pred. No. 1.2e-83; Mismatches 41; Indels 0; Gaps 0;
Matches 179; Conservative 31;

QY 1 KEVVLLDDEGIIYVHAHSDWSPVSHVPRKGGITVVPNDKDELIPORITTYGRMYIDFK 60
Db 826 KEILLLDAGVITYPISDSVWSPVHCVPKDKGMIVYKNEKDELIPRTITGHRMCIDYRK 885
QY 61 LNKATRKDHYPPLPFDHMLERLSKLTTHFCFLDGYSSFSQIPVAOSDOEKTTPCFFGFA 120
Db 886 LNAASRKHFPLPFDIDMLERLANHPYCCLDGYSGFQIPHPNDQKTEFTCPGTFA 945
QY 121 YRRMPFGLCNAPATFORCMAIFSNFCENIVEFMDDFSVYSSFPDCLSNDRYLQCK 180
Db 946 YRRMPFGLCNAPATFORCMTSIFSDLEIEMVEFMDDFSVYSSFPDCLSNDRYLQCK 1005
QY 181 DTNVLNCKEKFHVNNEGIVGKHISERGIEVDKAKYDAIDKMPYPTDIKIGIRSLGHG 240

```

Db 1006 ETNLVLMNEKCHFWREGIVLGRKISEEGIEVDKAKIDVMMOLOPPKTVKDIRSFLGHAG 1065
 QY 241 FYRRFIKDFPK 251
 Db 1066 FYRRFIKDFPK 1076

RESULT 4

Q9LHC0 PRELIMINARY; PRT; 897 AA.

AC Q9LHC0; PRELIMINARY; PRT; 897 AA.
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE RETROELEMENT POL. POLYPROTEIN-LIKE.
 CC Arabidopsis thaliana (Mouse-ear cress).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 RN NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Nakamura Y.;
 RT "Structural Analysis of Arabidopsis thaliana Chromosome 3. III."
 RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
 TRANSCRIPTASE).
 DR EMBL: AP002054; BAB02630.1; -
 DR InterPro: IPR001584; Rve.
 DR InterPro: IPR00477; RVTse.
 DR Pfam: PF00665; rve; 1.
 DR Pfam: PF0078; rvt; 1.
 KW Polypeptide: RNA-directed DNA polymerase.
 SQ SEQUENCE 897 AA; 103002 MW; 007AGA04B2C7CE18 CRC64;

Query Match 70.3%; Score 947; DB 10; Length 897;
 Best Local Similarity 71.3%; Pred. No. 3.2e-81;
 Matches 174; Conservative 29; Mismatches 41; Indels 0; Gaps 0;

QY 8 DEGTIYVHNSDWSVPVSHVPRKGGITVVPNDKDELIPORTITGYRMVIDEIKKATPK 67
 Db 66 DAGVIVPISDSWVSLVYCVPRKGGMTVANKEDILPRTITGHRMCIDYRKLNASK 125
 QY 68 DHPPLPFDIDMLERLANHRYCFLDGYSGFQPIHPNDQKTTFTCPYGFAYKRMFG 185
 Db 126 DHPPLPFDIDMLERLANHRYCFLDGYSGFQPIHPNDQKTTFTCPYGFAYKRMFG 185
 QY 128 ICNAPATFORCMAIFSNFCENIVEFMDPSVYSSFDCLSLNDRVLQRCRQTNVLIN 187
 Db 186 ICNAPATFORCMAIFSNFCENIVEFMDPSVYSSFDCLSLNDRVLQRCRQTNVLIN 245
 QY 188 GECCHWVNGIYLGHRKISRGIEVAKVADIKMYPPIIDKIGISFSGHGFYRRFK 247
 Db 246 WECCHWVNGIYLGHRKISRGIEVAKVADIKMYPPIIDKIGISFSGHGFYRRFK 305
 QY 248 DFTK 251
 Db 306 DFTK 309

RESULT 5

Q9SHM3 PRELIMINARY; PRT; 1799 AA.

AC Q9SHM3; PRELIMINARY; PRT; 1799 AA.
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE F7E2.17.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 RN NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RA Chao O., Shih P., Brooks S., Buehler E., Dunn P., Khan S., Kim C.,
 RA Walker M., Brooks S., Altafi H., Araujo R., Conn L., Conway A.B.,
 RA Gonzalez A., Hansen N.F., Hultzer L., Kremenetskaia I., Lenz C., Li J.,
 RA Liu S., Luros S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V.,
 RA Yu G., Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
 RT "Genomic sequence for Arabidopsis thaliana BAC F7P22 from chromosome
 I."
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
 TRANSCRIPTASE).
 DR EMBL: AC007534; BA624531.1; -
 DR InterPro: IPR001584; Rve.
 DR InterPro: IPR00477; RVTse.
 DR Pfam: PF00665; rve; 1.
 DR Pfam: PF0078; rvt; 1.
 KW RNA-directed DNA polymerase.
 SQ SEQUENCE 1799 AA; 204493 MW; 0E5BACC0930579FC CRC64;

Query Match 70.2%; Score 946; DB 10; Length 1799;
 Best Local Similarity 69.7%; Pred. No. 9.9e-81;
 Matches 175; Conservative 31; Mismatches 39; Indels 6; Gaps 1;

QY 1 KEVKKLIDEGITVHNSDWSVPVSHVPRKGGITVVPNDKDELIPORTITGYRMVIDEIKK 60
 Db 967 KEVKKLIDEGITVHNSDWSVPVSHVPRKGGITVVPNDKDELIPORTITGYRMVIDEIKK 1020
 QY 61 LNKATRKDHPPLPFDIDMLERLANHRYCFLDGYSGFQPIHPNDQKTTFTCPYGF 120
 Db 1021 LNKATRKDHPPLPFDIDMLERLANHRYCFLDGYSGFQPIHPNDQKTTFTCPYGF 1080
 QY 121 YRMPFGICNAPATFORCMAIFSNFCENIVEFMDPSVYSSFDCLSLNDRVLQRCR 180
 Db 1081 YRMPFGICNAPATFORCMAIFSNFCENIVEFMDPSVYSSFDCLSLNDRVLQRCR 1140
 QY 181 DTNLVLMNEKCHFWREGIVLGRKISEEGIEVDKAKIDVMMOLOPPKTVKDIRSFLGHAG 240
 Db 1141 DTNLVLMNEKCHFWREGIVLGRKISEEGIEVDKAKIDVMMOLOPPKTVKDIRSFLGHAG 1200
 QY 241 FYRRFIKDFPK 251
 Db 1201 FYRRFIKDFPK 1211

RESULT 6

Q9MOT8 PRELIMINARY; PRT; 724 AA.

AC Q9MOT8; PRELIMINARY; PRT; 724 AA.
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PURATIVE ATHILA TRANSPOSIN PROTEIN.
 GN ATG07660.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 RN NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project.
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
 TRANSCRIPTASE).
 DR EMBL: AL161506; CAB81136.1; -
 DR InterPro: IPR00477; RVTse.
 DR Pfam: PF0078; rvt; 1.
 KW RNA-directed DNA polymerase.
 SQ SEQUENCE 724 AA; 82390 MW; B80EB52DB35D5368 CRC64;

Query Match 62.4%; Score 841; DB 10; Length 724;

Best Local Similarity 66.8%; Pred. No. 2,8e-71;

Matches 155; Conservative 32; Mismatches 45; Indels 0; Gaps 0;

QY 1 KEVVLDEGIIYHVAHSDWSPVHSPKGGITVVPNDKDELIIPORITGYRMVIDPRK 60
 DB 367 KEILKLDAAGIYIPISDSTWEPVHCVPKKGMTVYKNEKDELIITRTITGHRVCIDYRK 426
 QY 61 LNKATRKDHVPLPFDHMLERLSKLTDFGSSSQIIPAOSDQKTFCTCPGTA 120
 DB 427 LNAASRKDHFPPTNQLNLNCLFDGYSGFQIIPNDQKTFCTCPGTA 486
 QY 121 YRRMPFGICNAPATFORCMAIFSNFCENIVEFMDPSVYSSFDCLSLNDRVLQCK 180
 DB 487 YRRMPFGICNAPATFORCMTISFDLTKMVEVFMDPSVYSSFDCLSLNDRVLQCK 546
 QY 181 DTNLVLNGECKHFVNNEGIVLGHKISERGIEVDKAKVDAIDKMPYPTDIKGI 232
 DB 547 ETNLVLNMECKYFMVKEGIVLGHKISEKGIEVDKEKIKVMQLODPKTVKDI 598

RESULT 7

092PTS PRELIMINARY; PRT; 841 AA.

AC 092PTS; PRELIMINARY; PRT; 841 AA.
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PUTATIVE POL. POLYPROTEIN WITH REVERSE TRANSCRIPTASE DOMAIN (PFAM: PF00078).
 GN F9B22.15.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucoids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., Vanaken S.E.,
 RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.,
 RA Carrera A.J., Cressy T.H., Buell C.R., Town C.D., Nierman W.C.,
 RA Frazer C.M., Venter J.C.;
 RT "Arabidopsis thaliana chromosome II BAC F9B22 genomic sequence."
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE TRANSCRIPTASE).
 CC EMBL: AC006528; AAD19780.1;
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF00078; RVT; 2.
 KW POLYprotein; RNA-directed DNA polymerase.
 SQ SEQUENCE 841 AA; 96367 MW; 788EFAF2477E3F0F CRC64;

Query Match 61.3%; Score 826; DB 10; Length 841;

Best Local Similarity 62.2%; Pred. No. 8,9e-70;

Matches 156; Conservative 30; Mismatches 37; Indels 28; Gaps 2;

QY 1 KEVVLDEGIIYHVAHSDWSPVHSPKGGITVVPNDKDELIIPORITGYRMVIDPRK 60
 DB 392 KEIMKLDAAGIYIPISDSTWEPVHCVPKKGMTVYKNEKDELIITRTITGHRVCIDYRK 451
 QY 61 LNKATRKDHVPLPFDHMLERLSKLTDFGSSSQIIPAOSDQKTFCTCPGTA 120
 DB 452 LNSATRKDNFPLSFIDOMLERLSNQPYCCFLDGYLGFGIILHPDQKTFCTCPGTA 511
 QY 121 YRRMPFGICNAPATFORCMAIFSNFCENIVEFMDPSVYSSFDCLSLNDRVLQCK 180
 DB 512 YRRMPFGICNAPATFORCMTISFDLTKMVEVFMDPSVYSSFDCLSLNDRVLQCK 556
 QY 181 DTNLVLNGECKHFVNNEGIVLGHKISERGIEVDKAKVDAIDKMPYPTDIKGI 240

DB 557 DKHLVLNMEKSHFVWBDGIVLGHKISEKGEVDRAKIE-----IMRFLGHAG 603

QY 241 FYRRFIKDPFK 251
 DB 604 FYRRFIKDPFK 614

RESULT 8

022103 PRELIMINARY; PRT; 407 AA.

AC 022103; PRELIMINARY; PRT; 407 AA.
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE REVERSE TRANSCRIPTASE-LIKE PROTEIN (FRAGMENT).
 DE Vicia faba (Broad bean).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucoids I; Fabales; Fabaceae; Papilionoideae; Vicia.
 OX NCBI_TaxID=3906;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kinoshita T., Mada H., Masaaki I., Shimazaki K.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE TRANSCRIPTASE).
 CC EMBL: AB007466; BAA22787.1;
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF00078; RVT; 1.
 KW RNA-directed DNA polymerase.
 FT NON_TPR 1
 SQ SEQUENCE 407 AA; 47607 MW; 010378486EF0748A CRC64;

Query Match 57.5%; Score 774; DB 10; Length 407;

Best Local Similarity 70.9%; Pred. No. 3e-65;

Matches 139; Conservative 29; Mismatches 28; Indels 0; Gaps 0;

QY 56 IDRKLKATRKDHVPLPFDHMLERLSKLTDFGSSSQIIPAOSDQKTFCTCP 115
 DB 3 IDRKLKATRKDHVPLPFDHMLERLADHEYCCFLDGYSGYNIAVPEQKTFCTCP 62
 QY 116 FGTFAYRRMPFGICNAPATFORCMAIFSNFCENIVEFMDPSVYSSFDCLSLNDRV 175
 DB 63 FGTFAYRRMPFGICNAPATFORCMTISFDLTKMVEVFMDPSVYSSFDCLSLNDRV 122
 QY 176 LORCKDTNLVLNGECKHFVNNEGIVLGHKISEKGIEVDKAKVDAIDKMPYPTDIKGI 235
 DB 123 LERCOESNLILNMECKHFVNREGIVLGHKISYKIGIEVDQAKIEVSLHPTNEKGI 182
 QY 236 LGHGFYRRFIKDPFK 251
 DB 183 LGHGFYRRFIKDPFK 198

RESULT 9

09C803 PRELIMINARY; PRT; 884 AA.

AC 09C803; PRELIMINARY; PRT; 884 AA.
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE POLYPROTEIN, PUTATIVE.
 GN T4121.10.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucoids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;

SQ SEQUENCE 565 AA; 65009 MW; 87797E91062B55AD CRC64;
 Query Match
 Best Local Similarity 44.6%; Pred. No. 2e-36;
 Matches 95; Conservative 17; Mismatches 37; Indels 64; Gaps 1;
 QY 18 SDWSPVHSPKKGITVVPNDKDELLPQRIITGYRWIDRKLKATRKDHYPLFDH 77
 Db 236 SDWSPVHSPKKGITVVPNDKDELLPQRIITGYRWIDRKLKATRKDHYPLFDH 77
 QY 78 MLERLSKLTHFCFLDGYSSFSQIPVPAOSDQKTFCTGFCGFAYRRMFGICNAPATFOR 137
 Db 251 -----FFQIHPDDQKTFCTGFCGFAYRRMFGICNAPATFOR 291
 QY 138 CMAIIFSCFENIVEFMDFSVYSSGSDCLSLMDLVLQCKDNTNLVNGKCHFWME 197
 Db 292 YMSIFTFYMEDFTEFVMDPFSYGFSEACLENLCKVLARCEKKNLLNMEKHFVYHE 351
 QY 198 GIVLGHKISERGTEVDKAKDAIDKMPYPTDIK 230
 Db 352 GIVLGHKISERGTEVDKAKDAIDKMPYPTDIK 230
 RESULT 12
 Q9VZ16 PRELIMINARY; PRT; 1062 AA.
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
 DE CG2485 PROTEIN.
 GN MIDLINE-JUMPER.POL OR CG2485 OR CG17399.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Butts K.C., Buxton M.R., Bock J., Brokstein P., Brotlier P.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferrar C., Ferrier A., Fleischmann W.,
 RA Flier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Gloder A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck M.,
 RA Hostalka M., Houson K.A., Howland T.J., Wei M.-H., Ibbegan C.,
 RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Liao X., Lele Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Maltsev B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svistkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Wolley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
 CC TRANSCRIPTASE).
 CC EMBL: AE003485; AF048015.1;
 DR FLYBASE: FBgn0040019; midline-jumper.vpol.
 DR InterPro: IPR001584; Rve.
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF00665; rve.1.
 DR Pfam: PF00078; rvt.1.
 KW RNA-directed DNA polymerase.
 SQ SEQUENCE 1062 AA; 123369 MW; 7BBCDA1E88BFE4B2 CRC64;
 Query Match
 Best Local Similarity 33.8%; Score 455; DB 5; Length 1062;
 Matches 96; Conservative 52; Mismatches 88; Indels 14; Gaps 3;
 QY 2 EVVNLDEGIIYVHSDWSPVHSPKKGITVVPNDKDELLPQRIITGYRWIDRKL 61
 Db 205 QINKLEQDILRH-SHSPWSAPVFLVPRK-----LDASNKK-----WRLVDRQL 250
 QY 62 NKATRKHYPLPFDHMLERLSKLTHTFCFLDGYSSFSQIPVPAOSDQKTFCTGFCGFAY 121
 Db 251 NDKTIKRYPMPTNELLDLGRAQYFALDLASGYHQIEVEPRDRSKTAFSAVGHEF 310
 QY 122 RRMFGICNAPATFORCMAIFSNFCENIVEFMDFSVYSSGSDCLSLMDLVLQCKD 181
 Db 311 IRMPFGICNAPATFORCMAIFSNFCENIVEFMDFSVYSSGSDCLSLMDLVLQCKD 370
 QY 182 TNLVNGKCHFWMEGIVLGHKISERGTEVDKAKDAIDKMPYPTDIKIRSFLLGHGF 241
 Db 371 ANKLQPDKSEFLKKELEYLGHIVTEKGVKPNKTIETIAFPMPTRKIRKISFLGLG 430
 QY 242 YRRFIDFKR 251
 Db 431 YRRFIDFKR 440
 RESULT 13
 Q9T0B7 PRELIMINARY; PRT; 1240 AA.
 AC Q9T0B7;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
 DE PUTATIVE REVERSE-TRANSCRIPTASE-LIKE PROTEIN.
 GN TAP9.40 OR ATAG10580.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RA Bevan M., Wedler E., Wambutt R., Bancroft I., Mewes H.W.,
 RA Mayer K.F.X., Schueller C.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Wedler H., Wedler E., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;
 RU Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
 CC TRANSCRIPTASE).
 DR EMBL: AL049523; CAB40024.1; -;
 DR EMBL: AL161517; CAB78181.1; -;
 DR InterPro: IPR001584; RVE.
 DR InterPro: IPR000477; RVTSE.
 DR Pfam: PF00665; rve; 1.
 DR Pfam: PF00078; rvt; 1.
 KW RNA-directed DNA polymerase.
 SQ SEQUENCE 1240 AA; 139488 MW; 40605C5ED73FED15 CRC64;

Query Match 33.7%; Score 453.5; DB 10; Length 1240;
 Best Local Similarity 39.4%; Pred. No. 3.2e-34;
 Matches 98; Conservative 41; Mismatches 91; Indels 19; Gaps 2;

QY 1 KEVVKLLDEGIYVHAHSDWVSPVSHVKKGGITVVPNDKDELIPQILITGYRNVDFRK 60
 DB 509 KQLDKLLGKGF-SPSTSPMGAPVLFVKKKG-----SFRCLIDYRE 549
 QY 61 LNKATRKDHVPLPFIDHMLERLSKLTFCFLDGYSSFSQIPVAOSDOEKTFTCPGTF 120
 DB 550 LNRVTYVKNRPLPRIDELDLQLRGATCFPSKIDLTSGYHQPDAADYRKTAFTRYGHFE 609
 QY 121 YRRMPGLCNAPATFORCMAIFSNFCENIVEVMDSEVYSGSSFDCLSNLDRLQRC 180
 DB 610 FVMPFGLTNPAPVFMRLMNSVFOEFLDEFYIFIDILVYSKSPREGEVILRSMKLR 669
 QY 181 DTNVLNGEKCHFMVNEGIVLGHISERGIENVDAKAYDAIDKMPYPTDIKIRSLFGHGF 240
 DB 670 EOKFPAKLSKCFQOREMGLGHIVSAGVSVDPEKIEAIKDPPTNATEIRSFLLWAG 729
 QY 241 FYRRFIKDF 249
 DB 730 YRRFVKG 738

RESULT 14
 065353 PRELIMINARY; PRT; 1084 AA.
 AC 065353;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ORF B.
 OS Autographa californica nuclear polyhedrosis virus (AcMNPV).
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 OC Nucleopolyhedrovirus.
 CC NCBI_TaxID=46015;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MUTANT FP-D;
 RX MEDLINE=90258898; PubMed=1692964;
 RA Friesen P.D., Nissen M.S.;
 RT "Gene organization and transcription of TED, a lepidopteran
 retron transposon integrated within the baculovirus genome.";
 RL Mol. Cell. Biol. 10:3067-3077(1990).
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
 CC TRANSCRIPTASE).
 DR EMBL: M32662; AAA92249.1; -;
 DR InterPro: IPR001584; RVE.
 DR InterPro: IPR000477; RVTSE.
 DR Pfam: PF00665; rve; 1.
 DR Pfam: PF00078; rvt; 1.
 KW RNA-directed DNA polymerase.
 SQ SEQUENCE 1084 AA; 125586 MW; 77EBD0063F83F296 CRC64;

Query Match 33.2%; Score 447; DB 12; Length 1084;
 Best Local Similarity 39.6%; Pred. No. 1.1e-33;
 Matches 99; Conservative 42; Mismatches 95; Indels 14; Gaps 3;

QY 2 EVVKLLDEGIYVHAHSDWVSPVSHVKKGGITVVPNDKDELIPQILITGYRNVDFRK 61
 DB 157 QLTMLDGLIT-RPDSAMSSPIWVVPK-----IDASKOK-----WRLVDFEKL 202
 QY 62 NKATRKDHVPLPFIDHMLERLSKLTFCFLDGYSSFSQIPVAOSDOEKTFTCPGTF 121
 DB 203 NEKITDKKPIPNISDVIDKLGKCOYFTTLDLASFYQVEMKPODISKTAFTVEGHFEF 262
 QY 122 RMPFGLCNAPATFORCMAIFSNFCENIVEVMDSEVYSGSSFDCLSNLDRLQRC 181
 DB 263 LRMPMGLKNSPSTFQVMDNVLRGLQNNICLVYLDIIIVSTSOEHLNLEVFQRLRE 322
 QY 182 TNLVNGEKCHFMVNEGIVLGHISERGIENVDAKAYDAIDKMPYPTDIKIRSLFGHGF 241
 DB 323 SNFKIOMDSSEPKLETFATVIGHIISDGIKRPMDKISAIQKLIKRTKEIKQFGLLG 382
 QY 242 YRRFIKDF 251
 DB 383 YRRFIKDF 392

RESULT 15
 091KU4 PRELIMINARY; PRT; 1504 AA.
 AC 091KU4;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL PROTEIN T32820.F.
 GN T32820.F.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 CC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Washu;
 RT "The A. thaliana Genome Sequencing Project.";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Wilson R.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
 CC TRANSCRIPTASE).
 CC -1- SIMILARITY: BELONGS TO ZN-FINGER CCHC TYPE FAMILY.
 DR EMBL: AF262041; AAF67363.1; -;
 DR InterPro: IPR001969; Asp-Protease.
 DR InterPro: IPR001584; RVE.
 DR InterPro: IPR000477; RVTSE.
 DR InterPro: IPR001878; Znf_CCHC.
 DR Pfam: PF00665; rve; 1.
 DR Pfam: PF00078; rvt; 1.
 DR Pfam: PF00098; zf-CCHC; 2.
 DR SMART: SM00343; Znf_C2HC; 2.
 DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_1.
 KW RNA-directed DNA polymerase; Zinc-finger.
 KW RNA-directed DNA polymerase.
 SQ SEQUENCE 1504 AA; 168176 MW; 1B1340C2FEF91588 CRC64;

Query Match 32.5%; Score 437.5; DB 10; Length 1504;
 Best Local Similarity 36.6%; Pred. No. 1.3e-32;
 Matches 96; Conservative 42; Mismatches 92; Indels 19; Gaps 2;

QY 1 KEVVKLLDEGIYVHAHSDWVSPVSHVKKGGITVVPNDKDELIPQILITGYRNVDFRK 60
 DB 636 KQLELLGKGF-RLNTSPMRTSVLFVKKKG-----SFRCLIDYRE 676
 QY 61 LNKATRKDHVPLPFIDHMLERLSKLTFCFLDGYSSFSQIPVAOSDOEKTFTCPGTF 120

Db 677 LNRVTVKKKYPIDELLDLGATCESKIDLTSGYHQIPIAEADVRRKTAFTRYGHPE 736
QY 121 YRRMPEGLCNAPATFORCMAIFSNFCENIVEVFMDDPSYGGSSFDCLSNDRYLORCK 180
Db 737 FVMPFGLTNAPAFMRIMNSVFOEFLDEFYIIFIDILVYSKSPEDHVLRRVMEKLR 796
QY 181 DTNIVLNGECKHFVNNEGIVLGHKISERGIEVDKAKYDAIDKMPYPTDIKGI RSFLGHGG 240
Db 797 EEKLEAFLSKCSFWQRRKMGFLGHIVSVEGVSVDPEKIEAIRDWPPTNATEIRSFGLAG 856
QY 241 FYRRFIKDF 249
Db 857 YRRFVKG 865

Search completed: April 1, 2002, 22:46:11
Job time: 1963 sec